

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

IN RE APPLICATION OF: Akira KINNO, et al.

GAU:

SERIAL NO: 10/659,308

EXAMINER:

FILED: September 11, 2003

FOR: IDENTIFIER GENERATING METHOD, IDENTITY DETERMINING METHOD, IDENTIFIER TRANSMITTING METHOD, IDENTIFIER GENERATING APPARATUS, IDENTITY DETERMINING APPARATUS, AND IDENTIFIER TRANSMITTING APPARATUS

REQUEST FOR PRIORITY

COMMISSIONER FOR PATENTS
ALEXANDRIA, VIRGINIA 22313

SIR:

- ☐ Full benefit of the filing date of U.S. Application Serial Number , filed , is claimed pursuant to the provisions of 35 U.S.C. §120.
- ☐ Full benefit of the filing date(s) of U.S. Provisional Application(s) is claimed pursuant to the provisions of 35 U.S.C. §119(e): Application No. Date Filed
- ☒ Applicants claim any right to priority from any earlier filed applications to which they may be entitled pursuant to the provisions of 35 U.S.C. §119, as noted below.

In the matter of the above-identified application for patent, notice is hereby given that the applicants claim as priority:

<u>COUNTRY</u>	<u>APPLICATION NUMBER</u>	<u>MONTH/DAY/YEAR</u>
JAPAN	2002-267274	September 12, 2002

Certified copies of the corresponding Convention Application(s)

- ☒ are submitted herewith
- ☐ will be submitted prior to payment of the Final Fee
- ☐ were filed in prior application Serial No. filed
- ☐ were submitted to the International Bureau in PCT Application Number
Receipt of the certified copies by the International Bureau in a timely manner under PCT Rule 17.1(a) has been acknowledged as evidenced by the attached PCT/IB/304.
- ☐ (A) Application Serial No.(s) were filed in prior application Serial No. filed ; and
- ☐ (B) Application Serial No.(s)
☐ are submitted herewith
☐ will be submitted prior to payment of the Final Fee

Respectfully Submitted,

OBLON, SPIVAK, McCLELLAND,
MAIER & NEUSTADT, P.C.

Masayasu Mori

Registration No. 47,301

Joseph A. Scafetta, Jr.
Registration No. 26, 803

Customer Number

22850

Tel. (703) 413-3000
Fax. (703) 413-2220
(OSMMN 05/03)

日 本 国 特 許 庁
JAPAN PATENT OFFICE

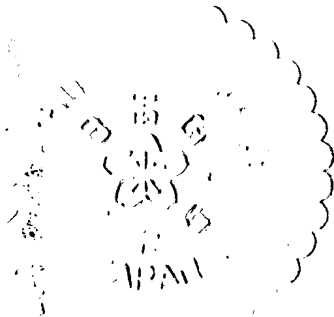
別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日
Date of Application: 2 0 0 2 年 9 月 1 2 日

出 願 番 号
Application Number: 特 願 2 0 0 2 - 2 6 7 2 7 4
[ST. 10/C]: [J P 2 0 0 2 - 2 6 7 2 7 4]

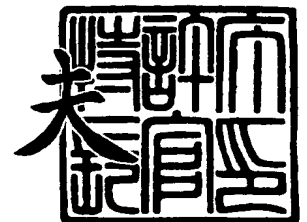
出 願 人
Applicant(s): 株式会社エヌ・ティ・ティ・ドコモ



2 0 0 3 年 9 月 1 1 日

特許庁長官
Commissioner,
Japan Patent Office

今 井 康



()

【書類名】 特許願

【整理番号】 14-0343

【提出日】 平成14年 9月12日

【あて先】 特許庁長官殿

【国際特許分類】 H09C 1/00

【発明者】

【住所又は居所】 東京都千代田区永田町二丁目 1 1 番 1 号 株式会社エヌ
・ ティ ・ ティ ・ ドコモ内

【氏名】 金野 晃

【発明者】

【住所又は居所】 東京都千代田区永田町二丁目 1 1 番 1 号 株式会社エヌ
・ ティ ・ ティ ・ ドコモ内

【氏名】 栄藤 稔

【発明者】

【住所又は居所】 東京都千代田区永田町二丁目 1 1 番 1 号 株式会社エヌ
・ ティ ・ ティ ・ ドコモ内

【氏名】 米本 佳史

【特許出願人】

【識別番号】 392026693

【氏名又は名称】 株式会社エヌ・ティ・ティ・ドコモ

【代理人】

【識別番号】 100088155

【弁理士】

【氏名又は名称】 長谷川 芳樹

【選任した代理人】

【識別番号】 100092657

【弁理士】

【氏名又は名称】 寺崎 史朗

【選任した代理人】

【識別番号】 100114270

【弁理士】

【氏名又は名称】 黒川 朋也

【選任した代理人】

【識別番号】 100108213

【弁理士】

【氏名又は名称】 阿部 豊隆

【選任した代理人】

【識別番号】 100113549

【弁理士】

【氏名又は名称】 鈴木 守

【手数料の表示】

【予納台帳番号】 014708

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 識別子生成方法、同一性判定方法、識別子送信方法、識別子生成装置、同一性判定装置および識別子送信装置

【特許請求の範囲】

【請求項 1】 文書データに対し、表記のゆらぎを正す規範化処理をほどこす規範化処理ステップと、

前記規範化処理ステップにおいて規範化処理が行われた前記文書データ全体もしくはその一部に基づいて、当該文書データもしくはその一部を一意に特定する識別子を生成する識別子生成ステップと、

を有することを特徴とする識別子生成方法。

【請求項 2】 前記規範化処理ステップは、型の定義が記述されている前記文書データのクラス定義ファイルを用いて、前記文書データに記述されている値を前記クラス定義ファイルによって定義された型に従って、表記の型を統一する型統一化処理ステップを有する、

ことを特徴とする請求項 1 に記載の識別子生成方法。

【請求項 3】 前記型統一化処理ステップにおいて、前記文書データのクラス定義ファイルに記述されている数値データの型定義にしたがって、前記文書データに記述されている数値データの精度を統一する、

ことを特徴とする請求項 2 に記載の識別子生成方法。

【請求項 4】 前記規範化処理ステップは、第 1 の部分データと第 2 の部分データとを所定の変換規則にしたがって文書データに変換する文書データ生成ステップを有する、

ことを特徴とする請求項 1 ～ 3 のいずれか 1 項に記載の識別子生成方法。

【請求項 5】 文書データに対し、表記のゆらぎを正す規範化処理をほどこす規範化処理ステップと、

前記規範化処理ステップにおいて規範化処理が行われた前記文書データ全体もしくはその一部に基づいて、当該文書データもしくはその一部を一意に特定する識別子を生成する識別子生成ステップと、

前記識別子生成ステップにおいて生成された識別子に基づいて、一の文書デー

タと他の文書データとで一致する部分があるか判定する同一性判定ステップと、
を有することを特徴とする同一性判定方法。

【請求項 6】 前記規範化処理ステップは、型の定義が記述されている前記文書データのクラス定義ファイルを用いて、前記文書データに記述されている値を前記クラス定義ファイルによって定義された型に従って、表記の型を統一する型統一化処理ステップを有する、

ことを特徴とする請求項 5 に記載の同一性判定方法。

【請求項 7】 前記型統一化処理ステップにおいて、前記文書データのクラス定義ファイルに記述されている数値データの型定義にしたがって、前記文書データに記述されている数値データの精度を統一する、

ことを特徴とする請求項 6 に記載の同一性判定方法。

【請求項 8】 前記規範化処理ステップは、第 1 の部分データと第 2 の部分データとを所定の変換規則にしたがって文書データに変換する文書データ生成ステップを有する、

ことを特徴とする請求項 5 ～ 7 のいずれか 1 項に記載の同一性判定方法。

【請求項 9】 前記識別子生成ステップにて生成された識別子を、前記文書データもしくは前記文書データに所定の処理を行った結果に関連付けてあらかじめキャッシュに格納する識別子格納ステップをさらに備え、

前記同一性判定ステップは、

前記識別子生成ステップにて生成された、同一性判定の対象となる文書データの識別子に基づいて前記キャッシュ内を検索し、前記識別子と同じ識別子が存在する場合は同一の文書データが存在すると判定し、前記識別子と同じ識別子が存在しない場合は同一の文書データが存在しないと判定する、

ことを特徴とする請求項 5 ～ 8 のいずれか 1 項に記載の同一性判定方法。

【請求項 10】 前記規範化処理ステップの前に、前記文書データ全体もしくはその一部に基づいて、当該文書データもしくはその一部を一意に特定する識別子を生成する第 2 の識別子生成ステップをさらに備え、

前記第 2 の識別子生成ステップにおいて生成された識別子に基づいて一の文書データと他の文書データとの同一性を判定し、同一と判定された場合は次の処理

ステップの処理をせずに終了し、同一でないと判定された場合は前記規範化処理ステップに処理を移行する、

ことを特徴とする請求項 5～9 のいずれか 1 項に記載の同一性判定方法。

【請求項 11】 文書データのエンコードデータ全体もしくはその一部に基づいて、当該文書データもしくはその一部を一意に特定する識別子を生成する識別子生成ステップと、

前記識別子生成ステップにおいて生成された識別子に基づいて、一の文書データと他の文書データとで一致する部分があるか判定する同一性判定ステップと、
を有することを特徴とする同一性判定方法。

【請求項 12】 前記同一性判定ステップにおいて前記文書データが同一であると判定された場合は、前記文書データのエンコードデータをデコードする処理をスキップする命令を出力する、

ことを特徴とする請求項 11 に記載の同一性判定方法。

【請求項 13】 文書データに対し、表記のゆらぎを正す規範化処理をほどこす規範化処理ステップと、

前記規範化処理ステップにおいて規範化処理が行われた前記文書データ全体もしくはその一部に基づいて、当該文書データもしくはその一部を一意に特定する識別子を生成する識別子生成ステップと、

前記識別子生成ステップにおいて生成された識別子を送信する識別子送信ステップと、

を有することを特徴とする識別子送信方法。

【請求項 14】 前記識別子送信ステップは、前記文書データ全体もしくはその一部を前記識別子によって置き換えたデータを生成する送信データ生成ステップを有する、

ことを特徴とする請求項 13 に記載の識別子送信方法。

【請求項 15】 前記送信データ生成ステップは、前記文書データに含まれる部分データを一意に特定する識別子と、前記部分データと文書データとの差分データと、によって記述した送信データを生成する、

ことを特徴とする請求項 14 に記載の識別子送信方法。

【請求項 16】 文書データに対し、表記のゆらぎを正す規範化処理をほどこす規範化処理手段と、

前記規範化処理手段によって規範化処理が行われた前記文書データ全体もしくはその一部に基づいて、当該文書データもしくはその一部を一意に特定する識別子を生成する識別子生成手段と、

を有することを特徴とする識別子生成装置。

【請求項 17】 前記規範化処理手段は、型の定義が記述されている前記文書データのクラス定義ファイルを用いて、前記文書データに記述されている値を前記クラス定義ファイルによって定義された型に従って、表記の型を統一する型統一化処理手段を有する、

ことを特徴とする請求項 16 に記載の識別子生成装置。

【請求項 18】 前記型統一化処理手段によって、前記文書データのクラス定義ファイルに記述されている数値データの型定義にしたがって、前記文書データに記述されている数値データの精度を統一する、

ことを特徴とする請求項 17 に記載の識別子生成装置。

【請求項 19】 前記規範化処理手段は、第 1 の部分データと第 2 の部分データとを所定の変換規則にしたがって文書データに変換する文書データ生成手段を有する、

ことを特徴とする請求項 16 ～ 18 のいずれか 1 項に記載の識別子生成装置。

【請求項 20】 文書データに対し、表記のゆらぎを正す規範化処理をほどこす規範化処理手段と、

前記規範化処理手段によって規範化処理が行われた前記文書データ全体もしくはその一部に基づいて、当該文書データもしくはその一部を一意に特定する識別子を生成する識別子生成手段と、

前記識別子生成手段によって生成された識別子に基づいて、一の文書データと他の文書データとで一致する部分があるか判定する同一性判定手段と、

を有することを特徴とする同一性判定装置。

【請求項 21】 前記規範化処理手段は、型の定義が記述されている前記文書データのクラス定義ファイルを用いて、前記文書データに記述されている値を

前記クラス定義ファイルによって定義された型に従って、表記の型を統一する型統一化処理手段を有する、

ことを特徴とする請求項 20 に記載の同一性判定装置。

【請求項 22】 前記型統一化処理手段は、前記文書データのクラス定義ファイルに記述されている数値データの型定義にしたがって、前記文書データに記述されている数値データの精度を統一する、

ことを特徴とする請求項 21 に記載の同一性判定装置。

【請求項 23】 前記規範化処理手段は、第 1 の部分データと第 2 の部分データとを所定の変換規則にしたがって文書データに変換する文書データ生成手段を有する、

ことを特徴とする請求項 20 ～ 22 のいずれか 1 項に記載の同一性判定装置。

【請求項 24】 前記識別子生成手段によって生成された識別子を、前記文書データもしくは前記文書データに所定の処理を行った結果に関連付けてあらかじめ格納したキャッシュをさらに備え、

前記同一性判定手段は、

前記識別子生成手段によって生成された、同一性判定の対象となる文書データの識別子に基づいて前記キャッシュ内を検索し、前記識別子と同じ識別子が存在する場合は同一の文書データが存在すると判定し、前記識別子と同じ識別子が存在しない場合は同一の文書データが存在しないと判定する、

ことを特徴とする請求項 20 ～ 23 のいずれか 1 項に記載の同一性判定装置。

【請求項 25】 前記規範化処理手段によって規範化処理を行う前に、前記文書データ全体もしくはその一部に基づいて、当該文書データもしくはその一部を一意に特定する識別子を生成する第 2 の識別子生成手段をさらに備え、

前記第 2 の識別子生成手段によって生成された識別子に基づいて一の文書データと他の文書データとの同一性を判定し、同一でないと判定された場合に前記規範化処理手段は文書データの規範化処理を行う、

ことを特徴とする請求項 20 ～ 24 のいずれか 1 項に記載の同一性判定装置。

【請求項 26】 文書データのエンコードデータ全体もしくはその一部に基づいて、当該文書データもしくはその一部を一意に特定する識別子を生成する識

別子生成手段と、

前記識別子生成手段によって生成された識別子に基づいて、一の文書データと他の文書データとで一致する部分があるか判定する同一性判定手段と、

を有することを特徴とする同一性判定装置。

【請求項 27】 前記同一性判定手段によって前記文書データが同一であると判定された場合は、前記文書データのエンコードデータをデコードする処理をスキップする命令を出力する、

ことを特徴とする請求項 26 に記載の同一性判定装置。

【請求項 28】 文書データに対し、表記のゆらぎを正す規範化処理をほどこす規範化処理手段と、

前記規範化処理手段によって規範化処理が行われた前記文書データ全体もしくはその一部に基づいて、当該文書データもしくはその一部を一意に特定する識別子を生成する識別子生成手段と、

前記識別子生成手段によって生成された識別子を送信する識別子送信手段と、を有することを特徴とする識別子送信装置。

【請求項 29】 前記識別子送信手段は、前記文書データ全体もしくはその一部を前記識別子によって置き換えたデータを生成する送信データ生成手段を有する、

ことを特徴とする請求項 28 に記載の識別子送信装置。

【請求項 30】 前記送信データ生成手段は、前記文書データに含まれる部分データを一意に特定する識別子と、前記部分データと文書データとの差分データと、によって記述した送信データを生成する、

ことを特徴とする請求項 29 に記載の識別子送信装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、表記は異なるが同一の意味を持つ文書データに対し、同一の識別子を生成する方法、その識別子を用いて同一性を判定する方法、生成された識別子を送信する方法と、識別子生成装置、同一性判定装置および識別子送信装置に関

する。

【0002】

【従来の技術】

XMLの普及に伴い、インターネットを代表とする広域ネットワーク上にあるさまざまなサービスの動的な接続を実現するアーキテクチャであるWebサービスが注目を浴びている。

【0003】

Webサービスでは、XML技術をネットワークプロトコル、サービスのインターフェース記述やコンテンツの管理などに利用することが考えられるが、XML文書は、タグによる記述を採用しているためファイルサイズが既存のHTTPのメッセージと比較して非常に大きくなる。そのため、ネットワークの負荷や、端末・サーバでの処理時間が長くなるといった問題がある。そこで、ネットワークの負荷軽減や処理の簡略化のために、XML文書から一意に生成される識別子にもとづいた処理を考える。

【0004】

識別子生成方法としては、XML文書を文字の羅列としてとらえ、一方向関数によって得られる結果値を識別子として生成している（例えば特許文献1を参照のこと）。しかし、XMLの仕様では、XML文書の記述を簡単化するため、空白がいくつあっても関係ない、クローズタグの省略形が存在する、コメントを記述することが可能、型の定義に沿っていればどのように記述してもかまわないなどといった、XML文書の記述者による表記の揺らぎによってXML処理に影響がおよばないよう柔軟に対応している。また、RDFの仕様では、文書データの構成要素がどのような順番にならんで記述されていても、文書データ全体としては同一の意味として取り扱うことが可能となっている。また、RDF仕様に従って定義されているCC/PPの仕様では、もともと定義してあるデフォルトデータをURIで指し示し、それに対する更新差分を記述することで、更新部分以外のデータの記述を省略が可能となっている。そのため、上述した従来の技術では、本来同一の意味をもったXML文書、RDF文書であったとしても、表記のゆらぎや型の違い、構成要素の並んでいる順番の違い、デフォルトデータおよび差分データによって記述されているな

ど、文字の羅列として文書を見ると異なった文書となる場合が十分に考えられる。すなわち、XML文書、RDF文書から一方向関数などを持ちいて識別子を生成することを考えると、同一の意味を持った文書であっても同一の識別子が生成されるとは限らない。

【0005】

【特許文献1】

特開 2001-282105号公報

【非特許文献1】

ISO/IEC 15938 Part 1 Systems Binary format-BiM

【0006】

【発明が解決しようとする課題】

本発明は、以上の問題を解決すべくなされたものであり、XML文書またはRDF文書が本来同一の意味を持った文書である場合、同一の識別子を生成する識別子生成方法、また、その識別子を用いて複数の文書データの同一性を判定する同一性判定方法、また、その識別子を送信する識別子送信方法と、それらの方法を実現可能な、識別子生成装置、同一性判定装置、及び識別子送信装置を提供することを目的とする。

【0007】

【課題を解決するための手段】

本発明に係る識別子生成方法（識別子生成装置）は、文書データに対し、表記のゆらぎを正す規範化処理をほどこす規範化処理ステップ（規範化処理手段）と、規範化処理ステップにおいて規範化処理が行われた文書データ全体もしくはその一部に基づいて、当該文書データもしくはその一部を一意に特定する識別子を生成する識別子生成ステップ（識別子生成手段）と、を有することを特徴とする。

【0008】

上記識別子生成方法（識別子生成装置）において、規範化処理ステップ（規範化処理手段）は、型の定義が記述されている文書データのクラス定義ファイルを用いて、文書データに記述されている値をクラス定義ファイルによって定義され

た型に従って、表記の型を統一する型統一化処理ステップ（型統一化処理手段）を有する、ことを特徴としても良い。

【0009】

上記識別子生成方法（識別子生成装置）は、型統一化処理ステップにおいて（型統一化処理手段によって）、文書データのクラス定義ファイルに記述されている数値データの型定義にしたがって、文書データに記述されている数値データの精度を統一する、ことを特徴としても良い。

【0010】

上記識別子生成方法（識別子生成装置）において、規範化処理ステップ（規範化処理手段）は、第1の部分データと第2の部分データとを所定の変換規則にしたがって文書データに変換する文書データ生成ステップ（文書データ生成手段）を有する、ことを特徴としても良い。

【0011】

本発明に係る同一性判定方法（同一性判定装置）は、文書データに対し、表記のゆらぎを正す規範化処理をほどこす規範化処理ステップ（規範化処理手段）と、規範化処理ステップにおいて（規範化処理手段によって）規範化処理が行われた文書データ全体もしくはその一部に基づいて、当該文書データもしくはその一部を一意に特定する識別子を生成する識別子生成ステップ（識別子生成手段）と、識別子生成ステップにおいて（識別子生成手段によって）生成された識別子に基づいて、一の文書データと他の文書データとで一致する部分があるか判定する同一性判定ステップ（同一性判定手段）と、を有することを特徴とする。

【0012】

上記同一性判定方法（同一性判定装置）において、規範化処理ステップ（規範化処理手段）は、型の定義が記述されている文書データのクラス定義ファイルを用いて、文書データに記述されている値をクラス定義ファイルによって定義された型に従って、表記の型を統一する型統一化処理ステップ（型統一化処理手段）を有する、ことを特徴としても良い。

【0013】

上記同一性判定方法（同一性判定装置）は、型統一化処理ステップにおいて（

手段によって)、文書データのクラス定義ファイルに記述されている数値データの型定義にしたがって、文書データに記述されている数値データの精度を統一する、ことを特徴としても良い。

【0014】

上記同一性判定方法(同一性判定装置)において、規範化処理ステップ(規範化処理手段)は、第1の部分データと第2の部分データとを所定の変換規則にしたがって文書データに変換する文書データ生成ステップ(文書データ生成手段)を有する、ことを特徴としても良い。

【0015】

上記同一性判定方法は、識別子生成ステップにて生成された識別子を、文書データもしくは文書データに所定の処理を行った結果に関連付けてあらかじめキャッシュに格納する識別子格納ステップをさらに備え、同一性判定ステップは、識別子生成ステップにて生成された、同一性判定の対象となる文書データの識別子に基づいてキャッシュ内を検索し、その識別子と同じ識別子が存在する場合は同一の文書データが存在すると判定し、識別子と同じ識別子が存在しない場合は同一の文書データが存在しないと判定する、ことを特徴としても良い。

【0016】

上記同一性判定装置は、識別子生成手段によって生成された識別子を、文書データもしくは文書データに所定の処理を行った結果に関連付けてあらかじめ格納したキャッシュをさらに備え、同一性判定手段は、識別子生成手段によって生成された、同一性判定の対象となる文書データの識別子に基づいてキャッシュ内を検索し、その識別子と同じ識別子が存在する場合は同一の文書データが存在すると判定し、識別子と同じ識別子が存在しない場合は同一の文書データが存在しないと判定する、ことを特徴としても良い。

【0017】

上記同一性判定方法は、規範化処理ステップの前に、文書データ全体もしくはその一部に基づいて、当該文書データもしくはその一部を一意に特定する識別子を生成する第2の識別子生成ステップをさらに備え、第2の識別子生成ステップにおいて生成された識別子に基づいて一の文書データと他の文書データとの同一

性を判定し、同一と判定された場合は次の処理ステップの処理をせずに終了し、同一でないと判定された場合は規範化処理ステップに処理を移行する、ことを特徴としても良い。

【0018】

上記同一性判定装置は、規範化処理手段によって規範化処理を行う前に、文書データ全体もしくはその一部に基づいて、当該文書データもしくはその一部を一意に特定する識別子を生成する第2の識別子生成手段をさらに備え、第2の識別子生成手段によって生成された識別子に基づいて一の文書データと他の文書データとの同一性を判定し、同一でないと判定された場合に規範化処理手段は文書データの規範化処理を行う、ことを特徴としても良い。

【0019】

また、本発明に係る同一性判定方法（同一性判定装置）は、文書データのエンコードデータ全体もしくはその一部に基づいて、当該文書データもしくはその一部を一意に特定する識別子を生成する識別子生成ステップ（識別子生成手段）と、識別子生成ステップにおいて（識別子生成手段によって）生成された識別子に基づいて、一の文書データと他の文書データとで一致する部分があるか判定する同一性判定ステップ（同一性判定手段）と、を有することを特徴とする。

【0020】

上記同一性判定方法（同一性判定装置）は、同一性判定ステップにおいて（同一性判定手段によって）文書データが同一であると判定された場合は、文書データのエンコードデータをデコードする処理をスキップする命令を出力する、ことを特徴としても良い。

【0021】

本発明に係る識別子送信方法（識別子送信装置）は、文書データに対し、表記のゆらぎを正す規範化処理をほどこす規範化処理ステップ（規範化処理手段）と、規範化処理ステップにおいて（規範化処理手段によって）規範化処理が行われた文書データ全体もしくはその一部に基づいて、当該文書データもしくはその一部を一意に特定する識別子を生成する識別子生成ステップ（識別子生成手段）と、識別子生成ステップにおいて（識別子生成手段によって）生成された識別子を

送信する識別子送信ステップ（識別子送信手段）と、を有することを特徴とする。

【0022】

上記識別子送信方法（識別子送信装置）において、識別子送信ステップ（識別子送信手段）は、文書データ全体もしくはその一部を識別子によって置き換えたデータを生成する送信データ生成ステップ（送信データ生成手段）を有する、ことを特徴としても良い。

【0023】

上記識別子送信方法（識別子送信装置）において、送信データ生成ステップ（送信データ生成手段）は、文書データに含まれる部分データを一意に特定する識別子と、部分データと文書データとの差分データと、によって記述した送信データを生成する、ことを特徴としても良い。

（作用）

【0024】

上記した課題を解決するために、本発明は、XML文書またはRDF文書の識別子を生成する前に、表記のゆらぎを回復する規範化処理を行う。XML-Canonicalizationに代表される規範化処理は、冗長な空白の削除、クローズタグ省略の解除など、XMLの仕様がゆるす表記のゆらぎを回復する処理である。規範化処理によって、同一の意味をもつXML文書またはRDF文書は同一の表記の文書となり、一方向関数に代表される文字の羅列から識別子を生成する関数によって、同一の識別子を生成することができる。

【0025】

また、本発明は、XML文書またはRDF文書のクラス定義ファイルを参照し、XML文書、RDF文書内に記述されているデータの型を統一化する。型の統一化処理によって、Double型やFloat型数字の精度などが統一化され、規範化処理同様、同一の意味をもつXML文書またはRDF文書は同一の表記の文書となり、一方向関数に代表される文字の羅列から識別子を生成する関数によって、同一の識別子を生成することができる。

【0026】

また、本発明は、CC/PP仕様であるデフォルトデータおよび差分データから文書データを記述できるように、複数の部分データを参照することで本来の文書データの意味を表す文書データに対し、部分データを取得し、ある変換規則に従ってその部分データを変換し、本来の文書データを生成してから識別子を生成する。識別子生成前に本来の文書データを生成することにより、同一の意味を持つXML文書、RDF文書は同一の表記の文書となり、一方向関数に代表される文字の羅列から識別子を生成する関数によって、同一の識別子を生成することができる。

【0027】

なお、本発明は、文書データの構成要素の順序を所定の規則にしたがって並べ替える処理を行った後に識別子を生成してもよい。

なお、本発明は、上述の識別子生成前の処理をそれぞれ備えることで、同一の意味をもつXML文書、RDF文書に対して同一の識別子を生成してもよい。

【0028】

また、本発明は、上述の生成手法によって生成される識別子を用いて、複数のXML文書またはRDF文書の同一性を判定することができる。なお、本発明に係る同一性判定方法（同一性判定装置）では、同一性判定ステップ（同一性判定手段）は、一の文書データと他の文書データとで一致する部分があるか判定する。すなわち、1）一の文書データの一部と他の文書データの一部が同一であるか、2）一の文書データと他の文書データの一部が同一であるか、3）一の文書データの一部と他の文書データが同一であるか、4）一の文書データと他の文書データが同一であるか、の判定を行うことができる。

【0029】

また、本発明は、規範化処理、型の統一化処理を段階的に配置し、それぞれの処理後に識別子を生成し、同一性を判定する。各同一性判定処理において同一と判定された場合は、次の段階に移行せずそのまま終了することで、同一性判定の処理時間を軽減できる。

【0030】

また、本発明は、XMLをエンコードしたデータ全体もしくはその一部から一意に生成した識別子をもちいることで、XMLの同一性を判定することができる。XML

のエンコードによって、同一の意味を持つ表記をあらかじめコード変換規則によって一意に定められたコードに割り当てるため（非特許文献 1 を参照のこと）、エンコードデータは、表記のゆれが正された状態になっている。すなわち、同一の意味を持つ XML 文書が同一のエンコードデータとして生成されるため、エンコードデータを文字の羅列として一方向関数などによって識別子を生成することによって、同一の意味をもつ XML 文書に対する同一性判定が可能となる。

【 0 0 3 1 】

本発明によれば、同一の意味をもつ複数の XML 文書、RDF 文書、またはその部分を同定することが可能となり、端末やサーバにおいて、過去に処理がされた XML 文書、RDF 文書またはその部分の処理が簡略化でき、処理時間を軽減することが可能となる。また、本発明によれば、XML 文書、RDF 文書またはその部分に対し一意に識別子を生成することができ、文書のデータ圧縮として利用できる。さらに同定することと同様に、端末やサーバにおける XML 文書、RDF 文書の処理が簡略化できる。

【 0 0 3 2 】

【発明の実施の形態】

〔第 1 実施形態〕

本発明の第 1 実施形態について以下に図面を参照しつつ説明する。図 1 は、本実施形態に係る識別子生成方法の動作を示すフロー図である。

【 0 0 3 3 】

同図のとおり、識別子生成方法は、識別子生成対象である文書データを取得する対象文書データ取得ステップ S101、対象文書データの表記のゆらぎを正す規範化処理ステップ S102、対象文書データ全体または選択した範囲から一意な識別子を生成する識別子生成ステップ S103 からなる。

【 0 0 3 4 】

図 1 1 は、入力された文書データの例として XML 文書の場合を示した図である。同図のとおり、入力された文書データは、無駄な空白や、クローズタグの省略形が用いられている。作成者による表記のゆれによって、生成される識別子が異ならないように、識別子生成の前に規範化を行う必要がある。

【0035】

図12は、図11で示した文書データに対し、XML-Canonicalizationの仕様にしたがって規範化をおこなったものの例である。同図のとおり、規範化処理後の文書データは、無駄な空白の削除、クローズタグの挿入がされ、作成者による表記のゆれがなくなっている。同図に示した文書データをもとに、識別子を生成する。

【0036】

識別子生成ステップS103では、規範化処理後の文書データ全体または選択した範囲から、一意な識別子を生成する。たとえば、ハッシュ関数などの一方向関数を持ち、ハッシュ値を識別子として生成する。ただし、識別子生成の関数は一方向関数である必要なく、一意な識別子が生成できる関数であればよい。

【0037】

図2は、規範化処理ステップS102において、対象文書データのクラス定義ファイルを確認し、型の統一化をする型の統一化処理ステップS201を付加した識別子生成方法の動作を示すフロー図である。

【0038】

図3は、型の統一化処理ステップS201の動作の詳細を示すフロー図である。同図のとおり、型の統一化処理は、対象文書データのクラス定義ファイルを取得するステップS301、クラス定義ファイルをもとに、対象文書データに記述されているすべてのデータの型を確認するデータ型確認ステップS302および対象文書データをクラス定義ファイルにしたがって変換をする文書データ変換ステップS303からなる。

【0039】

文書データ変換ステップS303は、データ型確認ステップS302において、クラス定義ファイルに記述されているデータタイプのとおりにデータを変換する。図14(a)及び図14(b)は対象文書データ変換前の文書データを示す図、図14(c)はクラス定義ファイルの例を示す図である。同図のとおり、クラス定義ファイルには、<ElementType name=“value” dt:type=“double” />と記述されており、valueの要素値の型がdouble型であることがわかる。また、図14(a

）に示す文書データ 1 および図 1 4 （b）に示す文書データ 2 には、それぞれ double 型である、12. 0、 12. 00 が value の要素値として記述してあることがわかる。どちらも型定義に沿った記述がなされているが、文字の羅列としてみた場合は異なる。

【0040】

図 1 5 （a）及び図 1 5 （b）は、対象文書データ変換後の文書データの例である。同図のとおり、double 型で定義されている value の要素値の精度を double 精度と同等にすることで、図 1 4 （a）及び図 1 4 （b）で示した文書データ 1、文書データ 2 の value 要素値 12. 0、12. 00 が等しくなり、型が統一化され、文字の羅列としてみた場合としても等しくなる。

【0041】

図 4 は、対象文書データがデフォルトデータおよび更新差分データからなりたっている場合、規範化処理ステップ S102 において、デフォルトデータおよび更新差分データから本来の文書データを生成する文書データ生成ステップ S401 を付加した識別子生成方法の動作を示すフロー図である。

【0042】

図 1 8 （a）は、デフォルトデータの例を示す図である。同図のとおり、デフォルトとして以下のとおり定義されている（第 10、11 行参照）。

```
<up:role>guest</up:role>
```

```
<up:age>16</up:age>
```

【0043】

図 1 8 （b）は、対象文書データとして RDF 文書の例を示す図である。同図のとおり、RDF 文書は 2 重定義がされており、デフォルトデータの URI（<ccpp:defaults rdf:resource= "UserProfileDefault"></ccpp:defaults>）と更新差分データ（<up:role>vip</up:role>）が記述されている（第 12、13 行参照）。この RDF 文書は、文書データ生成ステップ S401 において、本来の文書データに変換される。図 1 9 は、文書データ生成ステップ S401 によって変換された後の文書データである。同図は、更新差分データをデフォルトデータに上書きした場合を示す。同図のとおり、デフォルトデータの URI からデフォルトデータを取得し、更新差

分データによって上書きされ、`<up:role>guest</up:role>`が`<up:role>vip</up:role>`と変換されており（第12行参照）、本来の文書データとなっている。

【0044】

図26は、本実施形態に係る識別子生成方法をもちいて、対象文書データをその識別子とともにキャッシュに格納し、あとで識別子をもちいて対象文書データをキャッシュより取得することが可能な文書データ格納取得装置2700の構成を示すブロック図である。同図のとおり、文書データ格納取得装置2700は、対象文書データを取得する対象文書データ取得部2701、本実施形態に係る識別子生成方法が実装された識別子生成部2702、生成された識別子を対象文書データとともにキャッシュ2704に格納する識別子格納部2703、キャッシュ2704から取得したい文書データの識別子を取得する識別子取得部2705、識別子をもちいてキャッシュから文書データを取得する文書データ取得部2706からなる。

【0045】

図13は、識別子格納部2703によってキャッシュ2704に格納された文書データ識別子および文書データURIの例である。同図のように、入力された文書データを、識別子をもちいて管理することが可能となる。また、キャッシュにすでに格納されている識別子を格納しようとする際には、格納せずに文書データを廃棄することも可能となる。

【0046】

なお、図1、2、4で示す規範化処理ステップS102において、文書データ内の構成要素の順序を所定の規則にしたがって変更してもよい。例えば、図21のような複数の構成要素からなる文書データが入力された場合、`<rdf:Description rdf:about="xxxx">`のxxxxの部分がアルファベット順に整列するように構成要素の順序を並べ替える。図27は、構成要素の順序をアルファベット順に並べ替えた文書データである。RDFの仕様によれば、文書データ内の構成要素がどのような順序でならなくても文書データ全体でみれば同一の意味としてみなすことが可能である。すなわち、同一の意味をもつ文書データであっても、構成要素の順序が異なる場合が考えられる。そこで、構成要素の順序を、所定の規則にしたがっ

て変更したのちに、識別子を生成することで、構成要素の順序は異なるが同一の意味をもつ文書データに対して、同一の識別子を生成することが可能となる。

【0047】

[第2実施形態]

本発明の第2実施形態について以下に図面を参照しつつ説明する。図5は、本実施形態に係る同一性判定方法の動作を示すフロー図である。

【0048】

同図のとおり、同一性判定方法は、同一性判定対象である文書データを取得する対象文書データ取得ステップS501、対象文書データの表記のゆらぎを正す規範化処理ステップS502、対象文書データ全体または選択した範囲から一意な識別子を生成する識別子生成ステップS503、および、識別子生成ステップS503で生成された識別子をもとに複数の文書データの同一性を判定する同一性判定ステップS504からなる。

【0049】

規範化処理前の文書データおよび規範化処理後の文書データの例を図11、図12に示す。同図のとおり、規範化処理によって表記のゆらぎが回復される。

【0050】

図6は、規範化処理ステップS502において、対象文書データのクラス定義ファイルを確認し、型の統一化をする型の統一化処理ステップS601を付加した同一性判定方法の動作を示すフロー図である。

【0051】

図3は型の統一化処理ステップS601の動作の詳細を示すフロー図である。動作の詳細は第1実施形態と同じである。

【0052】

図14(a)及び図14(b)は型の統一化処理前の文書データ1、文書データ2をそれぞれ示す図、図14(c)はクラス定義ファイルの例を示す図、図15(a)及び図15(b)は型の統一化処理後の文書データ1、文書データ2の例をそれぞれ示す図である。同図のとおり、表記の異なっていたvalueの値が型の統一化ステップS601によって同一の表記となる。

【0053】

図7は、対象文書データがデフォルトデータおよび更新差分データからなりたっている場合、規範化処理ステップS502において、デフォルトデータおよび更新差分データから本来の文書データを生成する文書データ生成ステップS701を付加した同一性判定方法の動作を示すフロー図である。

【0054】

図18(a)及び図18(b)は、デフォルトデータおよび対象文書データの例、図19は文書データ生成ステップS701によって生成される変換後の文書データの例である。動作の詳細は第1実施形態と同じである。

【0055】

図8は、同一性判定ステップS504の動作の詳細を示したフロー図である。同図のとおり、同一性判定ステップS504は、同一性を判定する対象文書データから生成された識別子を取得する識別子取得ステップS801、識別子をキャッシュに格納する識別子格納ステップS803、キャッシュから当該識別子を検索し、検索成功の場合は同一の文書データが存在していると判断し、検索失敗の場合は同一の文書データが存在していないと判断する同一性判定ステップS802からなる。キャッシュに当該識別子が存在していない場合は、識別子を識別子格納ステップS803にわたし、当該識別子をキャッシュに格納する。

【0056】

図10は、本実施形態に係る同一性判定方法によって、入力される文書データがすでに処理されているかどうかを判定する同一性判定装置1101の構成を示すブロック図である。同図のとおり、同一性判定装置1101は対象文書データを取得する対象文書データ取得部1102、対象文書データから識別子を生成する識別子生成部1103、生成された識別子とキャッシュ1110をもちいて、すでに処理されているかどうかを判定する同一性判定部1108および、同一性判定部によってすでに処理されていないことが判定された場合は、次の同一性判定のために識別子をキャッシュ1110に格納する識別子格納部1109からなる。また、識別子生成部1103は、デフォルトデータおよび更新差分データから本来の文書データを生成する文書データ生成部1104、表記のゆらぎを回

復する規範化処理を行う規範化処理部 1105、クラス定義ファイルをもちいて、データの型統一化処理を行う型統一化処理部 1106 と、対象文書データから一意に識別子を生成する識別子生成処理部 1107 からなる。

【0057】

図 16 は、本実施形態に係る同一性判定方法をもちいることによって、すでに処理がされた文書データが入力された場合は、文書データの項書き換え処理をスキップする項書き換え処理装置の構成を示したブロック図である。同図のとおり、本実施形態に係る同一性判定方法が実装された同一性判定部 1701、項書き換え規則文書にしたがって文書データの項書き換えをおこなう項書き換え部 1702、項書き換え部 1702 にて生成された変換後の文書データを同一性判定部 1701 にて生成される文書データの識別子とともに格納してある変換後文書データキャッシュ 1703 からなる。

【0058】

同一性判定部 1701 は、文書データが入力されると、識別子を生成する。生成された識別子と変換後文書データキャッシュ 1703 をもちいて、すでに項書き換えが行われたことのある文書データかどうかを判定する。すでに項書き換えが行われたことのある文書データである場合は、項書き換え部 1702 の処理をスキップし、変換後文書データキャッシュ内に存在している当該識別子で取得できる変換後の文書データを出力する。本発明によって一般的に処理時間のかかる項書き換え処理をスキップすることが可能となるため高速な処理が可能となる。

【0059】

[第3実施形態]

本発明の第3実施形態について以下に図面を参照しつつ説明する。図9は、本実施形態に係る同一性判定方法の動作を示すフロー図である。

【0060】

同図のとおり、同一性判定方法は、同一性判定対象である文書データのエンコードデータを取得するエンコードデータ取得ステップS901、取得したエンコードデータ全体もしくはその一部から識別子を生成する識別子生成ステップS902、および識別子生成ステップS902で生成された識別子をもとに複数の文書データの同

一性を判定する同一性判定ステップS903からなる。

【0061】

XMLのエンコードによって、同一の意味を持つ表記は、あらかじめコード変換規則によって一意に定められたコードに割り当てられるため、エンコードデータは、表記のゆれが正された状態になっている。すなわち、同一の意味を持つXML文書が同一のエンコードデータとして生成されるため、エンコードデータを文字の羅列として一方向関数などによって識別子を生成することによって、同一の意味をもつXML文書に対する同一性判定が可能となる。

【0062】

図25は、同一の意味を持つが、文字の羅列としてみた場合異なる文書である文書データ1、文書データ2に対して、バイトコード表に従ってエンコードをほどこすことで、同一のエンコードデータが生成されることを示した図である。同図のとおり、改行が挿入されているだけで、同一の意味を持つ文書データ1、文書データ2に対し、エンコードをほどこすと、同一のエンコードデータが得られる。このエンコードデータをもとに識別子を生成することによって、同一の意味をもつ文書データに対して同一の識別子を生成することが可能となる。

【0063】

図16において、入力する文書データとして、XML文書をデータ圧縮したエンコードデータを入力とする。エンコードデータが入力された場合、処理をするにはデコードすることが必須となるが、本実施形態に係る同一性判定方法をもちいることで、エンコードデータの状態のまま同一性判定が可能となるため、すでに処理がすすんでいる文書データに冠してはデコード処理もスキップすることが可能となる。

【0064】

[第4実施形態]

本発明の第4実施形態について以下に図面を参照しつつ説明する。図17は、本実施形態に係る識別子送信方法の動作を示すフロー図である。

【0065】

同図のとおり、識別子送信方法は、対象となる文書データを取得する対象文書

データ取得ステップS1801、対象文書データに対して、規範化処理、型の統一化処理などを行い、文書データ全体もしくはその一部の識別子を生成する識別子生成ステップS1802、文書データ全体もしくはその一部を生成された識別子と置き換える識別子置換処理ステップS1803、識別子で代替される文書データまたはその一部をデフォルトとし、それとの差分データを文書データに付加する差分データ付加ステップS1804、以上の処理によって生成された文書データを送信する送信ステップS1805からなる。

【0066】

図20は、本実施形態に係る識別子送信方法を用いて、識別子つき文書データを送信する識別子送信装置2101の構成を示したブロック図である。同図のとおり、識別子送信装置2101は、対象文書データを取得する対象文書データ取得部1102、規範化処理、型統一処理などを行い、識別子を生成する識別子生成部1103及び対象文書に対し、識別子付加処理を行って送信する識別子送信部2102から構成される。

【0067】

識別子送信部2102は、文書データ全体もしくはその一部を識別子生成部1103によって生成された識別子と置き換える識別子置換処理部2103、識別子で代替される文書データ又はその一部をデフォルトとし、それとの差分データを文書データに付加する差分データ付加部2104、及び以上の処理によって生成された文書データを送信する送信部2105から構成される。

【0068】

図21は、対象文書データの例である。同図のとおり対象文書データは、<ccpp:component>…</ccpp:component>という構成要素が3つ記述されている。対象文書データは、識別子生成ステップS1802において、それぞれの構成要素について識別子が生成される。

【0069】

図22は、本実施形態に係る識別子送信方法によって送信される文書データの例として、図21に示した文書データの一部の記述（構成要素1、構成要素2）を識別子に置き換え、さらに、追加データとして構成要素3を記述した部分識別

子つき文書データの例である。同図のとおり、構成要素 1、構成要素 2 が、識別子におきかわることで、文書データ全体のデータ量が削減できる。

【0070】

図 23 は、本実施形態に係る識別子送信方法によって送信される部分識別子つき文書データをもちいたサービスシステム例を示した図である。同図のとおり、サービスシステムは、本実施形態に係る識別子送信方法によって送信される部分識別子つき文書データを送信する端末 2400 と、部分識別子つき文書データを受信し、部分識別子を本来の構成要素に展開して本来の文書データを生成し送信するプロキシ 2401 および、文書データを受信し、サービスを行うサーバ 2402 から構成される。プロキシ 2401 は、部分識別子と本来の構成要素とが関連付けられて格納されたデータベース 2403 と接続されている。端末 2400 からは、本来サーバに送信すべき文書データのうち、すでにプロキシを介して送信したことがある構成要素を識別子に置き換えて送信することが可能となるため、ネットワーク負荷のかからない文書データ送信方法として用いることが可能となる。

【0071】

[第 5 実施形態]

本発明の第 5 実施形態について以下に図面を参照しつつ説明する。図 24 は、本実施形態に係る段階的同一性判定方法の動作を示すフロー図である。

【0072】

同図のとおり、段階的同一性判定方法は、入力された文書データからそのまま識別子生成し、同一性の判定をする第 1 判定ステップ S2501、規範化処理を行ってから識別子を生成し、同一性の判定をする第 2 判定ステップ S2502、クラス定義ファイルをもちいて型の統一化を行ってから識別子を生成し、同一性を判定する第 3 判定ステップ S2503、および判定結果を出力する結果出力ステップ S2504 からなる。

【0073】

第 1 判定ステップ S2501 で同一と認められなかった場合は、第 2 判定ステップ S2502 に処理をうつす。第 2 判定ステップ S2502 においても、第 1 判定ステップ

S2501と同様、同一と認められなかった場合は、第3判定ステップS2503に処理をうつす。第1判定ステップS2501、第2判定ステップS2502とも、同一と認められた場合は、次以降の処理をせずに結果出力ステップS2504において判定結果を出力する。第3判定ステップは、同一性の判定を行った結果を結果出力ステップS2504にわたし、判定結果を出力する。

【0074】

【発明の効果】

本発明によれば、同一の意味を持つ文書データ、またはその一部に対して、同一の識別子を生成することが可能となるため、同一の意味をもつ文書データを同定することが可能となる。

【0075】

また、本発明によれば端末やサーバにおいて、過去に処理がされた文書データが入力された場合は、同一性判定方法によって処理をスキップすることが可能となり、処理時間を軽減することができる。

【0076】

また、本発明によれば、文書データをエンコードデータの状態のまま同一性を判定することが可能となるため、端末やサーバにおけるデコード処理もふくめた処理のスキップが可能となり、処理時間を軽減することが可能となる。

【0077】

また、本発明によれば、同一の意味をもつ文書データ全体もしくはその一部から生成した識別子を文書データ全体もしくはその一部とおきかえた文書データを生成することが可能となるため、文書データのデータ量を削減することが可能となり、たとえば文書データを送信する際に、すでに送信済みであった部分を識別子に置き換えて送信することで、ネットワーク負荷を軽減することが可能となる。

【図面の簡単な説明】

【図1】

実施形態に係る識別子生成方法の動作を示すフロー図である。

【図2】



実施形態に係る識別子生成方法の動作を示すフロー図である。

【図 3】

型の統一化処理ステップS201の動作の詳細を示すフロー図である

【図 4】

実施形態に係る識別子生成方法の動作を示すフロー図である。

【図 5】

実施形態に係る同一性判定方法の動作を示すフロー図である。

【図 6】

実施形態に係る同一性判定方法の動作を示すフロー図である。

【図 7】

実施形態に係る同一性判定方法の動作を示すフロー図である。

【図 8】

実施形態に係る同一性判定方法の動作を示すフロー図である。

【図 9】

実施形態に係る同一性判定方法の動作を示すフロー図である。

【図 10】

実施形態に係る同一性判定装置の構成を示すブロック図である。

【図 11】

入力された文書データの例としてXML文書の場合を示した図である。

【図 12】

図 11 で示した文書データに対し、XML-Canonicalizationの仕様にしたがって規範化をおこなったものの例を示す図である。

【図 13】

キャッシュに格納された文書データ識別子および文書データURIの例を示す図である。

【図 14】

(a) 及び (b) は対象文書データ変換前の文書データを示す図、(c) はクラス定義ファイルの例を示す図である。

【図 15】



(a) 及び (b) は、対象文書データ変換後の文書データの例を示す図である。

【図 1 6】

項書き換え処理装置の構成を示すブロック図である。

【図 1 7】

実施形態に係る識別子送信方法の動作を示すフロー図である。

【図 1 8】

(a) はデフォルトデータの例を示す図、(b) は、対象文書データとして R D F 文書の例を示す図である。

【図 1 9】

文書データ生成ステップ S401 によって変換された後の文書データである。

【図 2 0】

実施形態に係る識別送信装置の構成を示すブロック図である。

【図 2 1】

複数の構成要素からなる文書データの例を示す図である。

【図 2 2】

本実施形態に係る識別子送信方法によって送信される文書データの例を示す図である。

【図 2 3】

実施形態に係る識別子送信方法によって送信される部分識別子つき文書データをもちいたサービスシステム例を示す図である。

【図 2 4】

実施形態に係る段階的同一性判定方法の動作を示すフロー図である。

【図 2 5】

エンコードデータについて説明する図である。

【図 2 6】

文書データ格納取得装置の構成を示すブロック図である。

【図 2 7】

構成要素の順序をアルファベット順に並べ替えた文書データを示す図である。

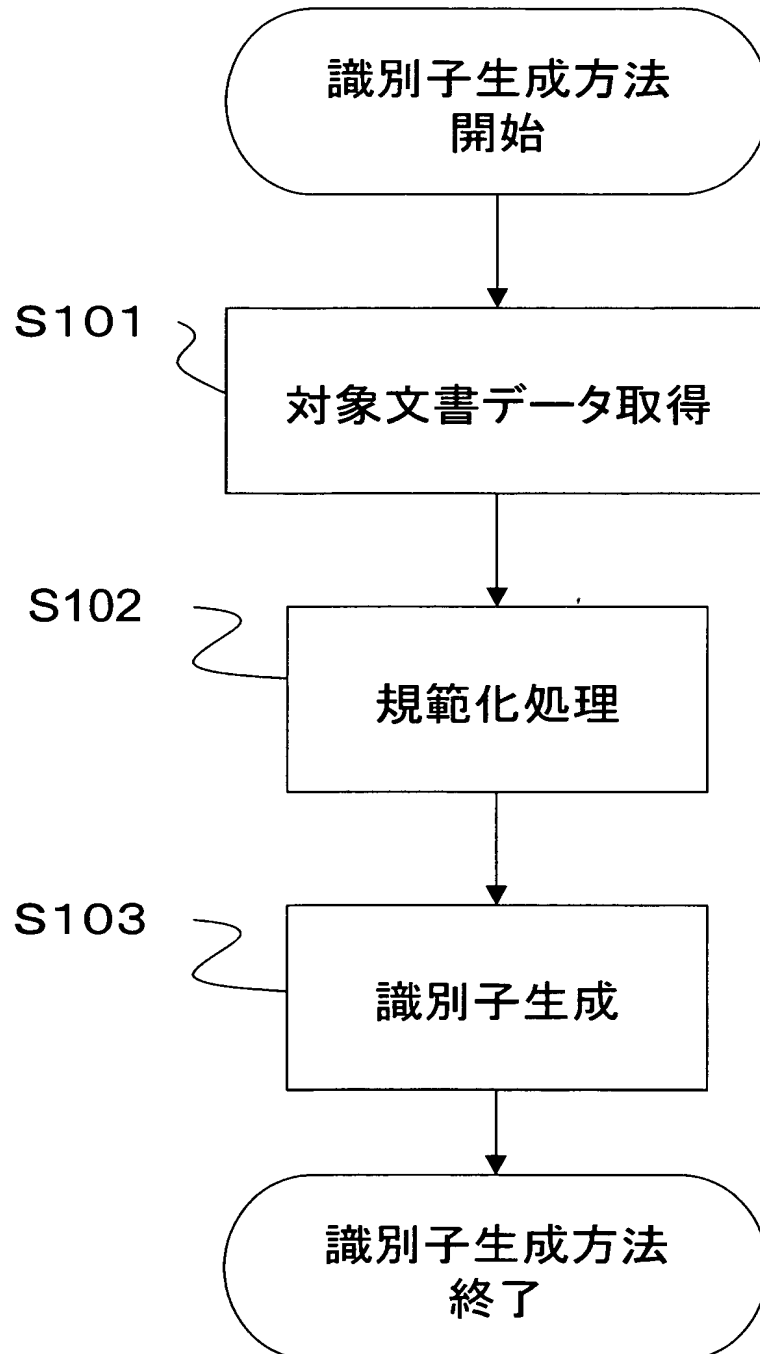
**【符号の説明】**

1101…同一性判定装置、1102…対象文書データ取得部1103…識別子生成部、1108…同一性判定部、1109…識別子格納部、1110…キャッシュ、1701…同一性判定部、1702…項書き換え部、1703…変換後文書データキャッシュ、2101…識別子送信装置、2102…識別子送信部、2103…識別子置換処理部、2104…差分データ付加部、2400…端末、2401…プロキシ、2402…サーバ、2403…データベース、2700…文書データ格納取得装置、2701…対象文書データ取得部、2702…識別子生成部、2703…識別子格納部、2704…キャッシュ、2705…識別子取得部、2706…文書データ取得部。

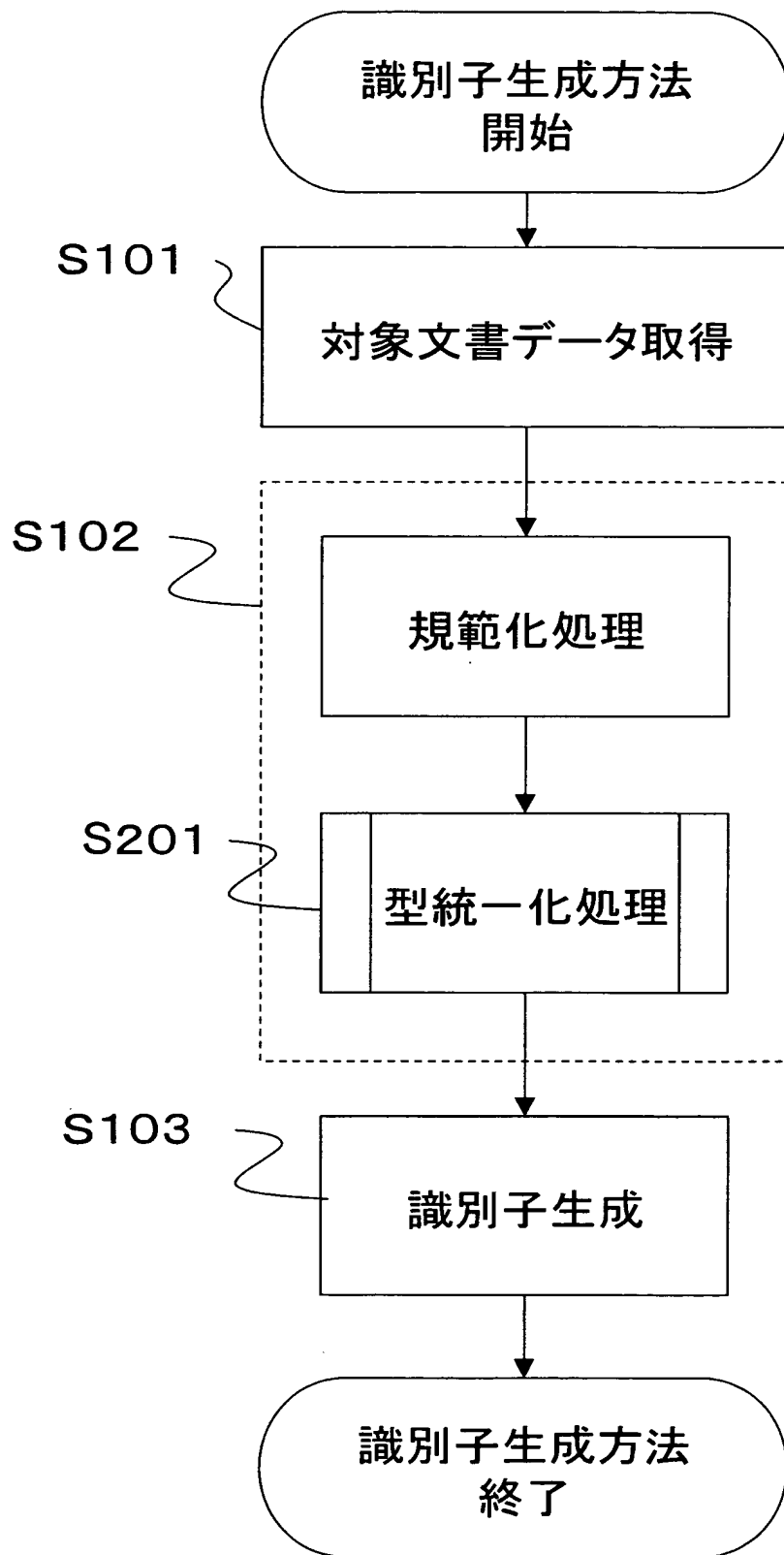
【書類名】

図面

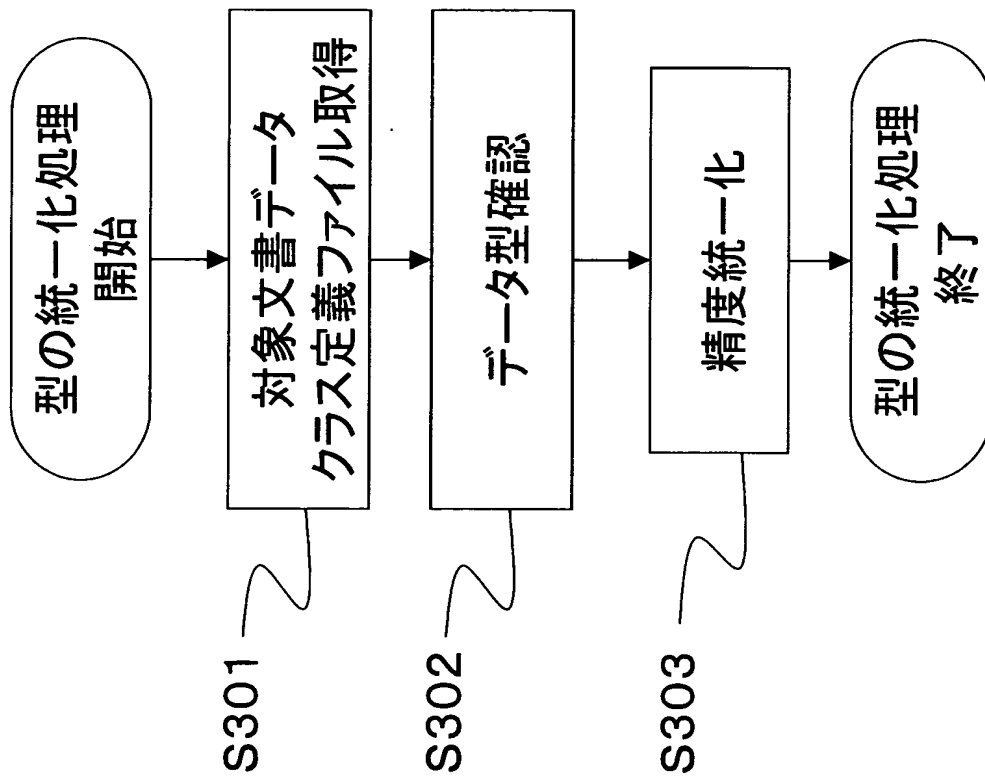
【図 1】



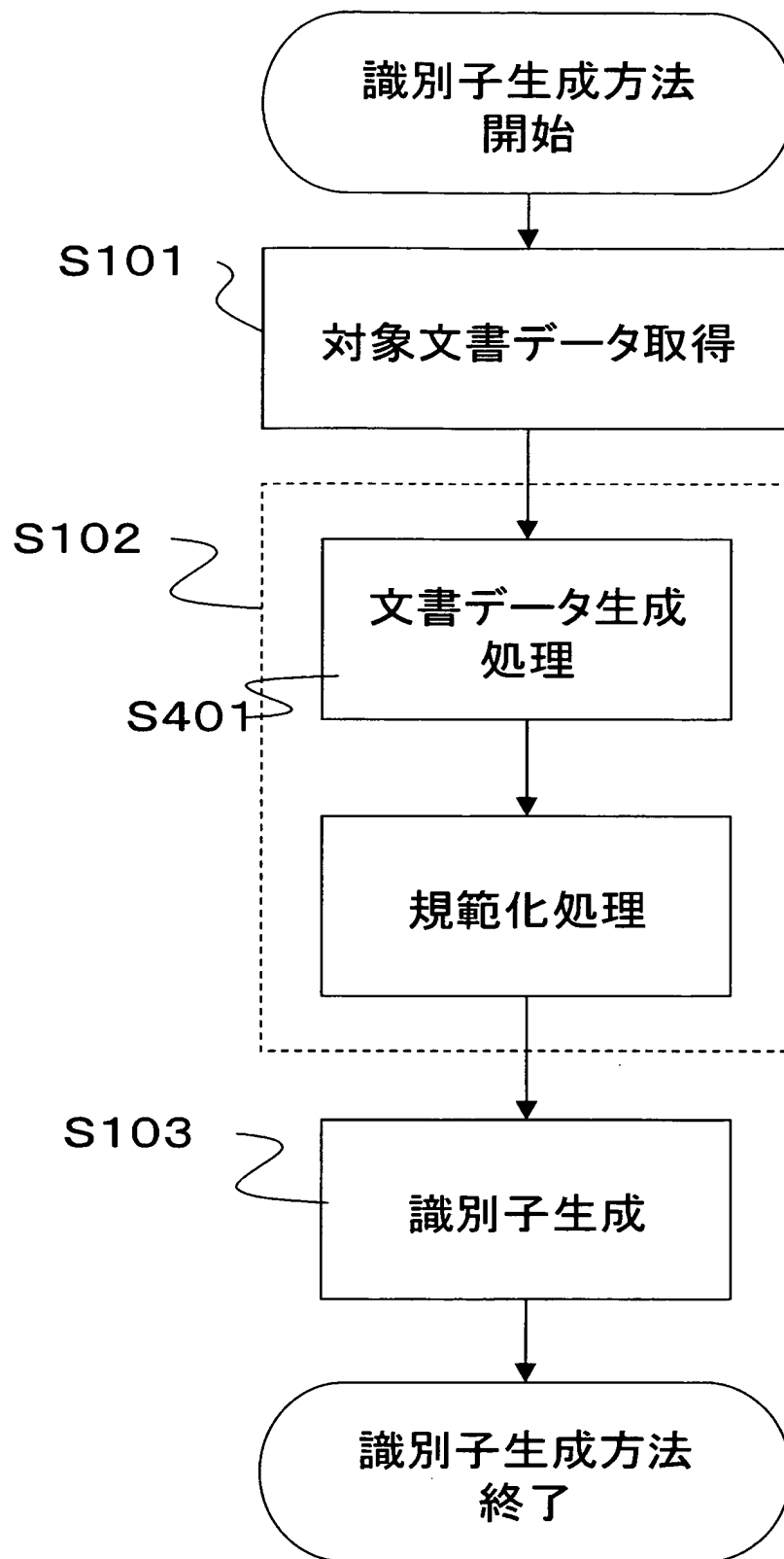
【図 2】



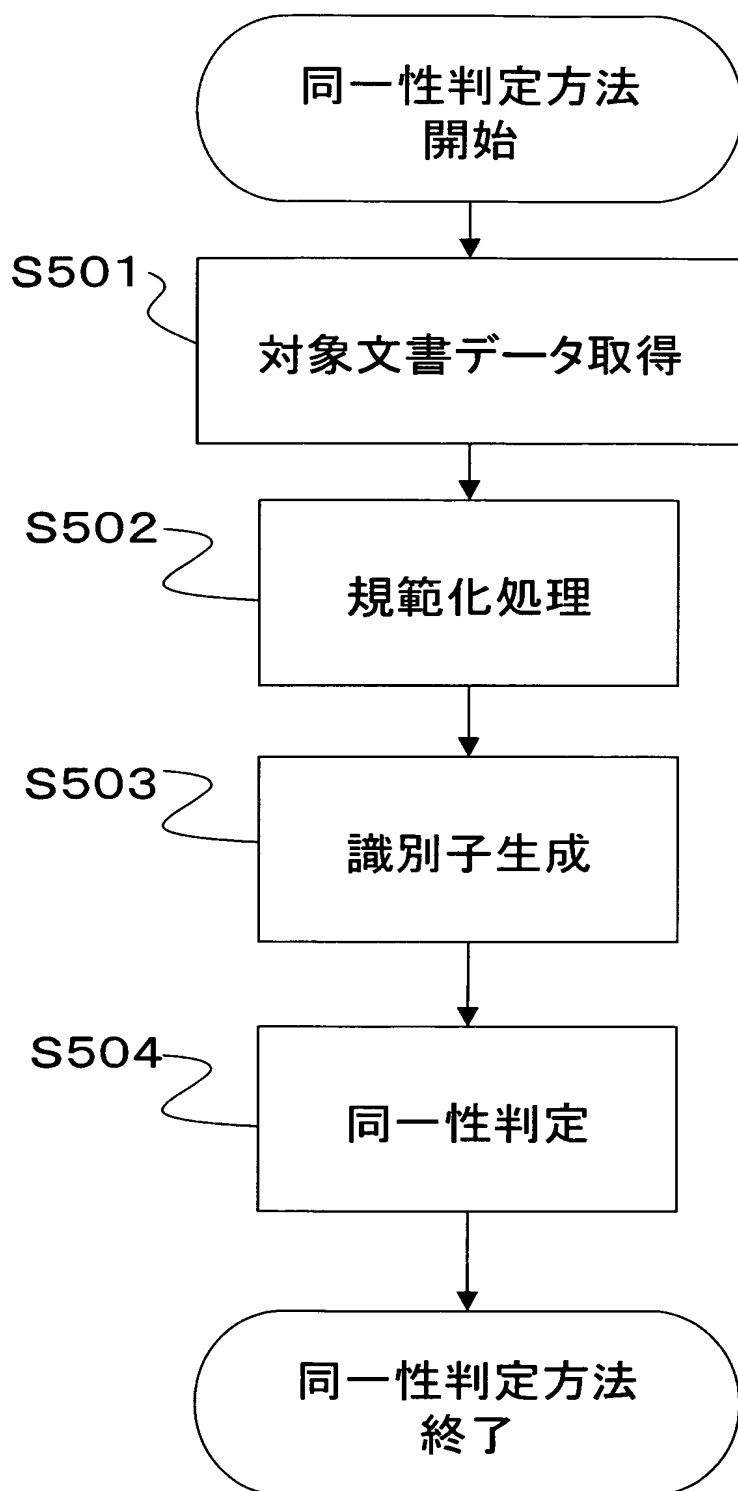
【図 3】



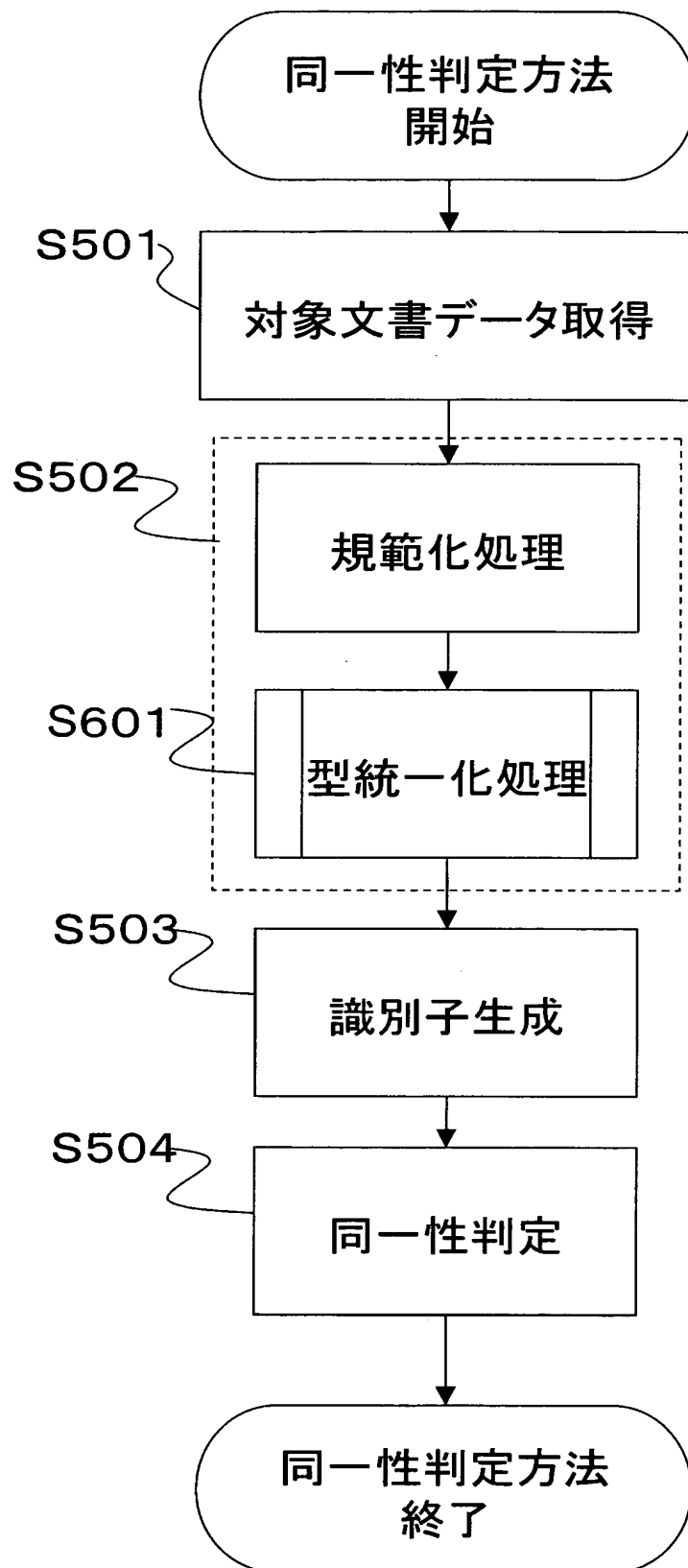
【図 4】



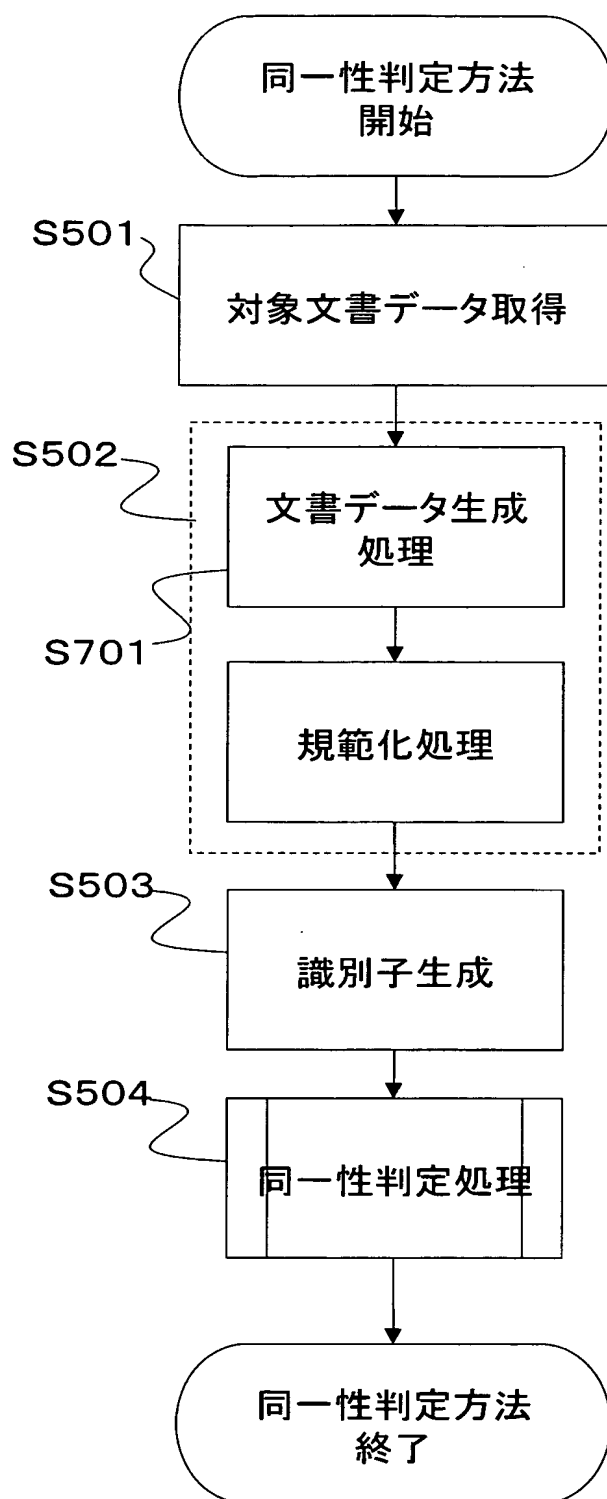
【図 5】



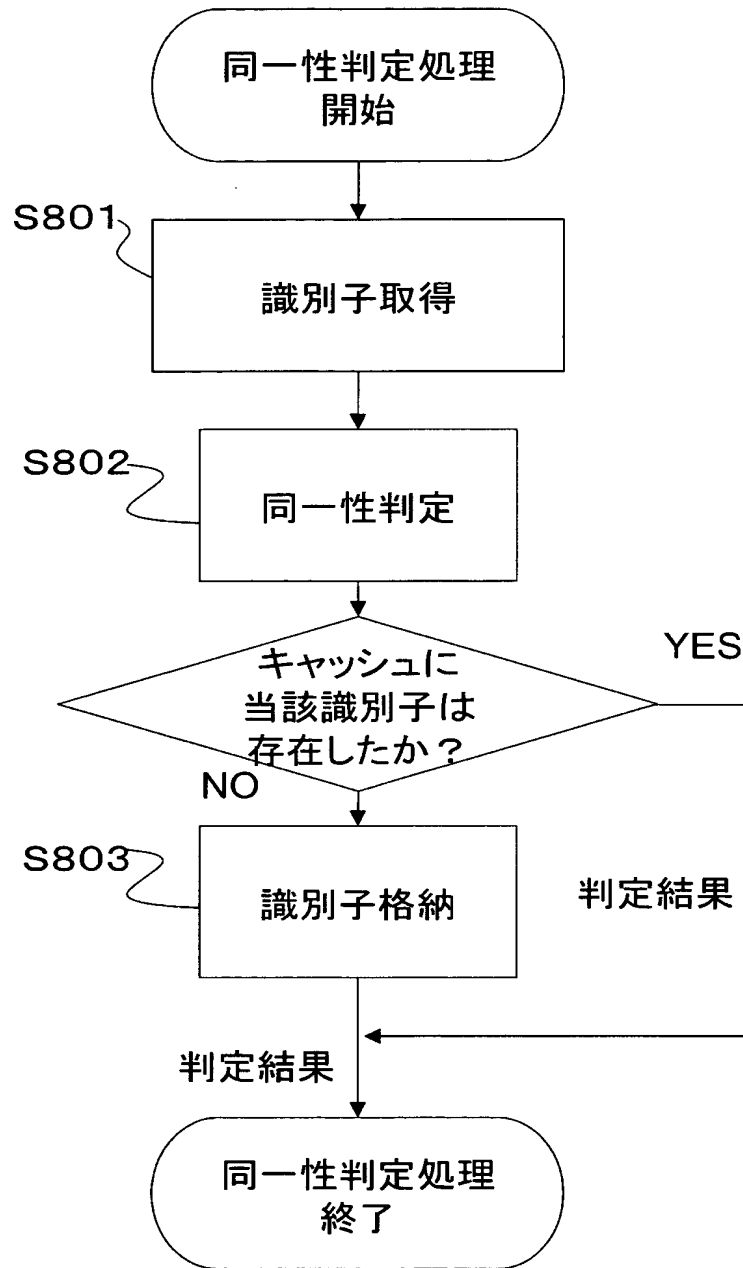
【図 6】



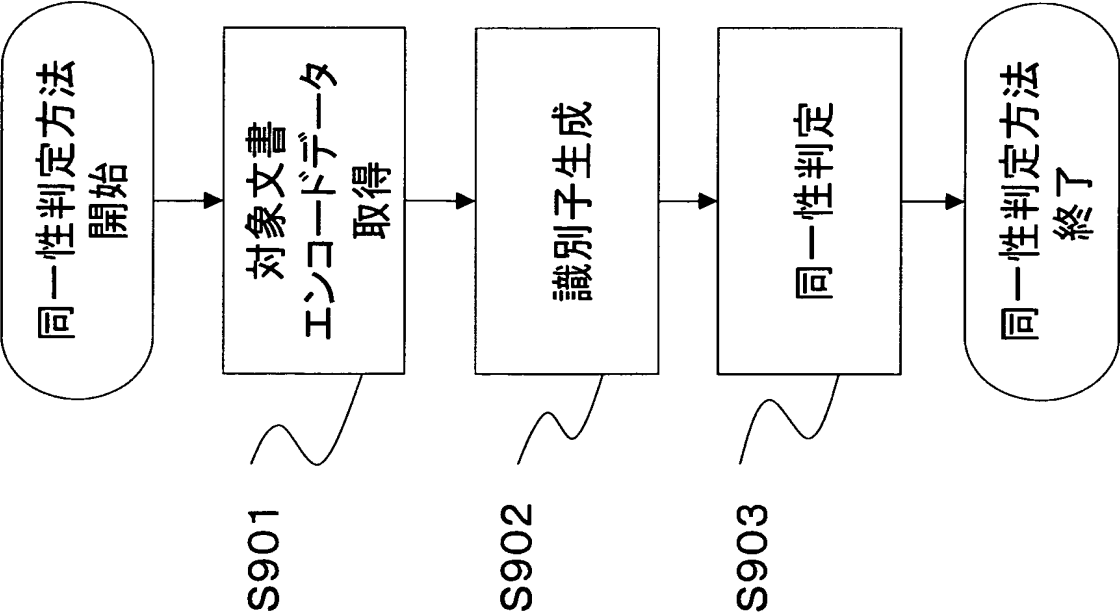
【図 7】



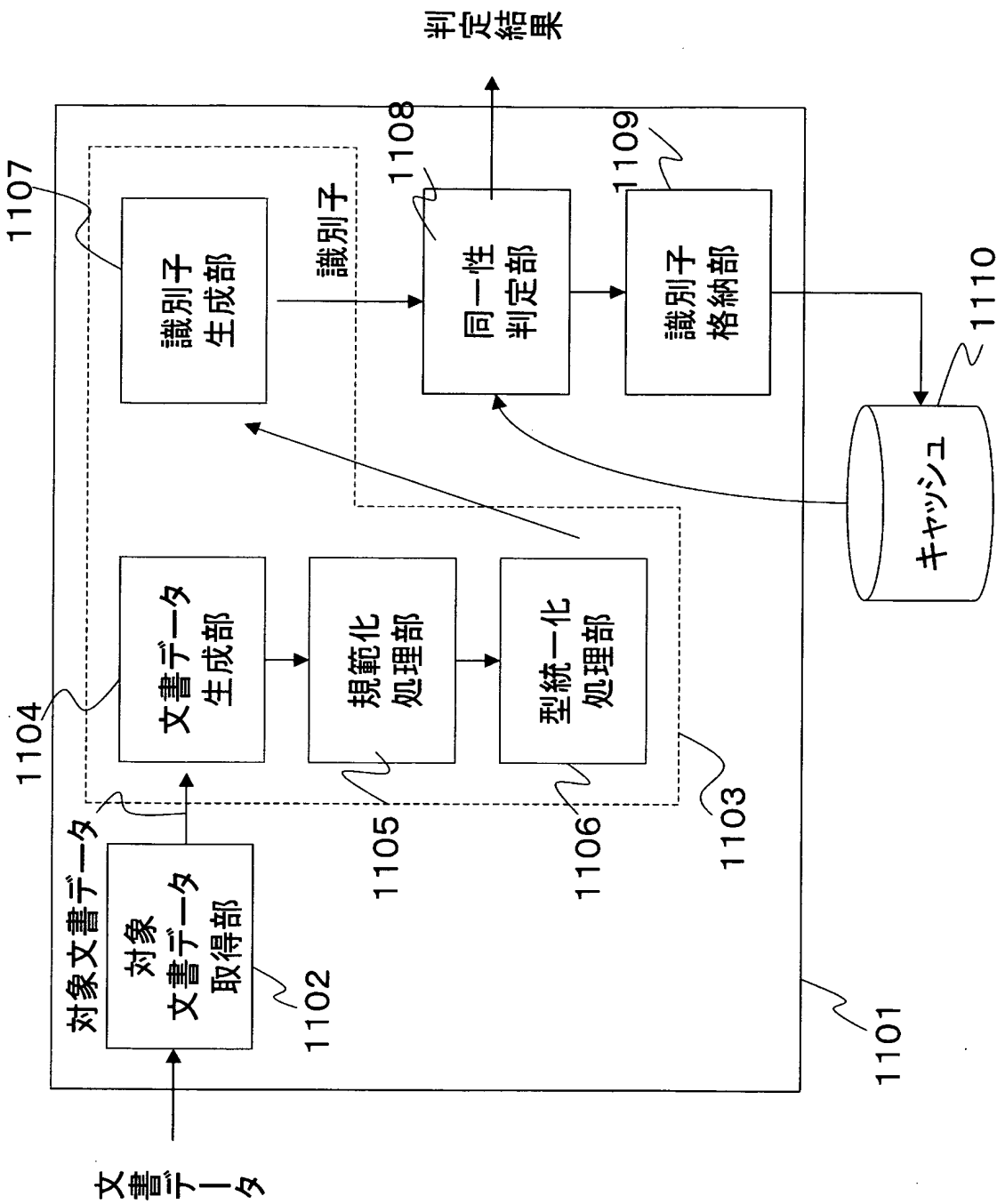
【図 8】



【図 9】



【図 10】



【図 11】

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns"
  xmlns:ccpp="http://www.w3.org/2000/07/04-ccpp"
  xmlns:prf="http://www.wapforum.org/profiles/UAPROF/ccppschema-20010330"
  xmlns:pss5="http://www.3gpp.org/profiles/PSS/ccppschema-PSS5"
  xmlns:up="http://www.nttdocomo.co.jp/profiles/UP/ccppschema-UP"
  xmlns:upf="http://www.nttdocomo.co.jp/profiles/UPF/ccppschema-UPF">
  <rdf:Description rdf:about="http://www.bar.com/Phones/Phone007">
    <ccpp:component>
      <rdf:Description ID="UserProfile">
        <rdf:type rdf:resource="http://www.nttdocomo.co.jp/Profiles/UP/
          ccppschema-UP#UserProfile"/>
        <up:role> guest </up:role>
        <up:age>16</up:age>
      </rdf:Description>
    </ccpp:component>
  </rdf:Description>
</rdf:RDF>
```

【図 12】

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns"
  xmlns:ccpp="http://www.w3.org/2000/07/04-ccpp"
  xmlns:prf="http://www.wapforum.org/profiles/UAPROF/ccppschema-20010330"
  xmlns:pss5="http://www.3gpp.org/profiles/PSS/ccppschema-PSS5"
  xmlns:up="http://www.nttdocomo.co.jp/profiles/UP/ccppschema-UP"
  xmlns:upf="http://www.nttdocomo.co.jp/profiles/UPF/ccppschema-UPF">
  <rdf:Description rdf:about="http://www.bar.com/Phones/Phone007">
    <ccpp:component>
      <rdf:Description ID="UserProfile">
        <rdf:type rdf:resource="http://www.nttdocomo.co.jp/Profiles/UP/
          ccppschema-UP#UserProfile"></rdf:type>
        <up:role>guest</up:role>
        <up:age>16</up:age>
      </rdf:Description>
    </ccpp:component>
  </rdf:Description>
</rdf:RDF>
```

【図 1 3】

文書データ識別子	文書データURI
:	:
141325132	www.foo.co.jp/sheet141325132.xml
141325146	www.foo.co.jp/sheet141325146.xml
141325199	www.foo.co.jp/sheet141325199.xml
:	:

【図 1 4】

文書データ1

```
<?xml version="1.0"?>
<data>
  :
  <value> 12.0
  </value>
  :
</data>
```

(a)

文書データ2

```
<?xml version="1.0"?>
<data>
  :
  <value> 12.00
  </value>
  :
</data>
```

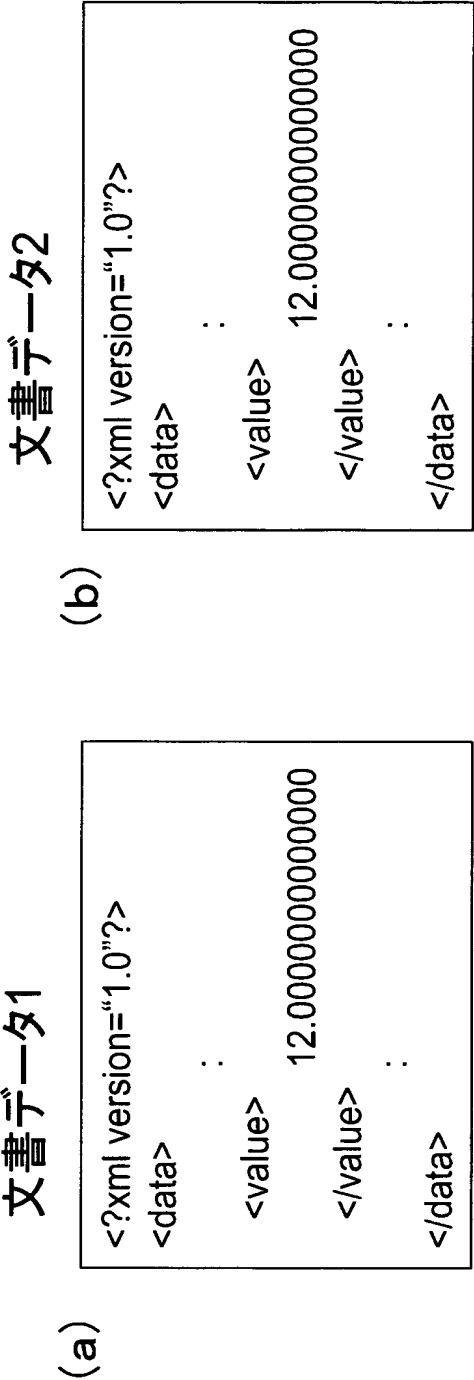
(b)

クラス定義ファイル

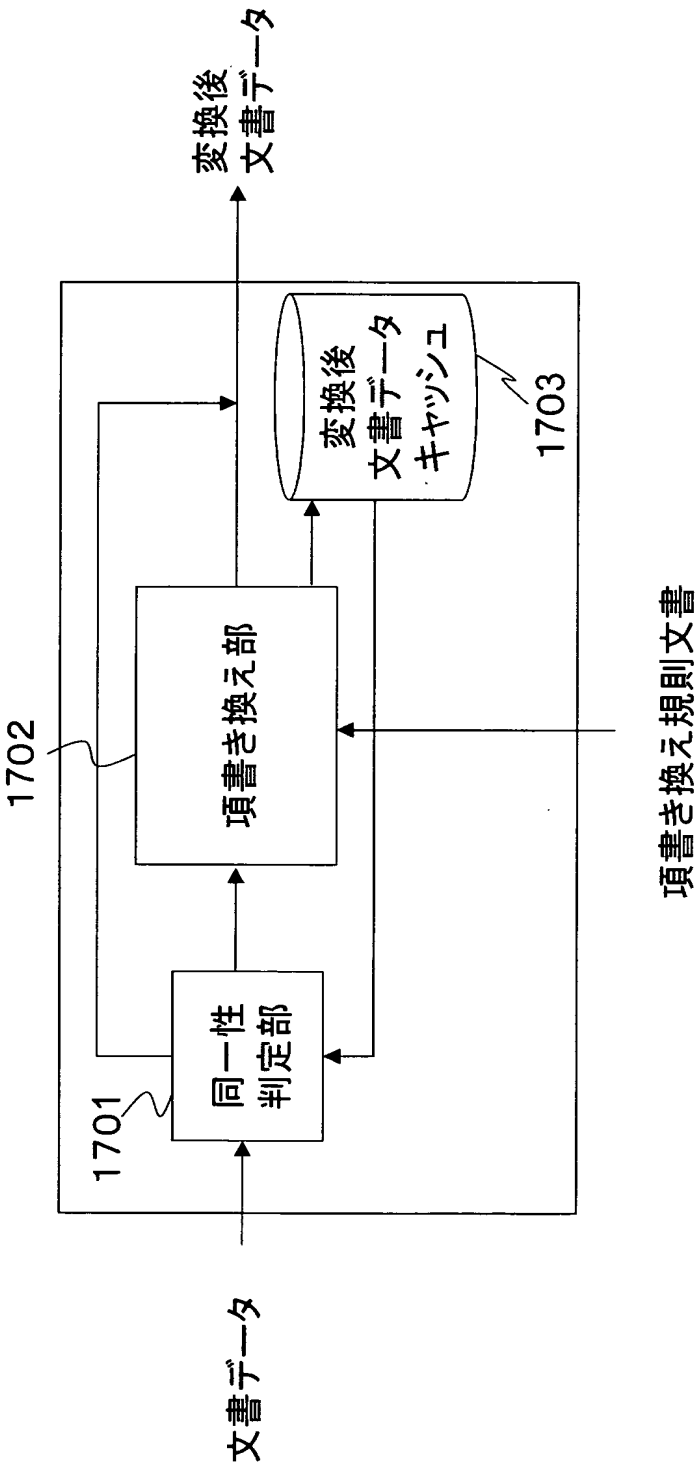
```
<Schema name="myschema"
  xmlns="urn:schemas-microsoft-com:xml-data"
  xmlns:dt="urn:schemas-microsoft-com:datatypes">
  :
  <ElementType name="value" dt:type="double"/>
  :
</Schema>
```

(c)

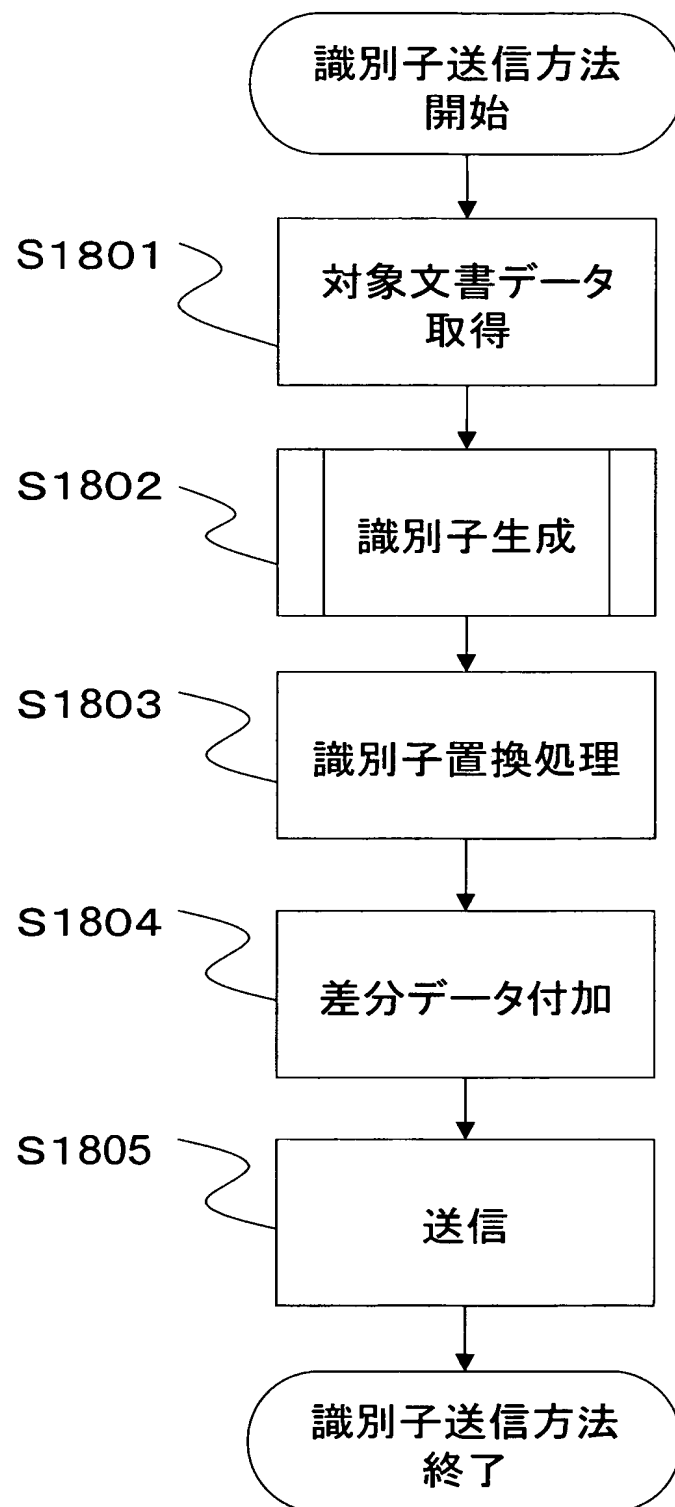
【図 1 5】



【図 16】



【図 17】



【図 18】

(a) デフォルト

```

<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22 -rdf-syntax-ns"
  xmlns:ccpp="http://www.w3.org/2000/07/04 -ccpp"
  xmlns:prf="http://www.wapforum.org/profiles/UAPROF/ccppschem -20010330"
  xmlns:pss5="http://www.3gpp.org/profiles/PSS/ccppschem -PSS5"
  xmlns:up="http://www.nttdocomo.co.jp/profiles/UP/ccppschem -UP"
  xmlns:upf="http://www.nttdocomo.co.jp/profiles/UPF/ccppschem -UPF">
  <rdf:Description rdf:about="UserProfileDefault">
    <rdf:type rdf:resource="UserProfilePlatform"></rdf:type>
    <up:role>guest</up:role>
    <up:age>16</up:age>
  </rdf:Description>
</rdf:RDF>

```

(b) 対象文書データ

```

<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22 -rdf-syntax-ns"
  xmlns:ccpp="http://www.w3.org/2000/07/04 -ccpp"
  xmlns:prf="http://www.wapforum.org/profiles/UAPROF/ccppschem -20010330"
  xmlns:pss5="http://www.3gpp.org/profiles/PSS/ccppschem -PSS5"
  xmlns:up="http://www.nttdocomo.co.jp/profiles/UP/ccppschem -UP"
  xmlns:upf="http://www.nttdocomo.co.jp/profiles/UPF/ccppschem -UPF">
  <rdf:Description rdf:about="MyProfile">
    <ccpp:component>
      <rdf:Description ID="UserProfile">
        <rdf:type rdf:resource="UserProfilePlatform"></rdf:type>
        <ccpp:defaults rdf:resource="UserProfileDefault"></ccpp:defaults>
        <up:role>vip</up:role>
      </rdf:Description>
    </ccpp:component>
  </rdf:Description>
</rdf:RDF>

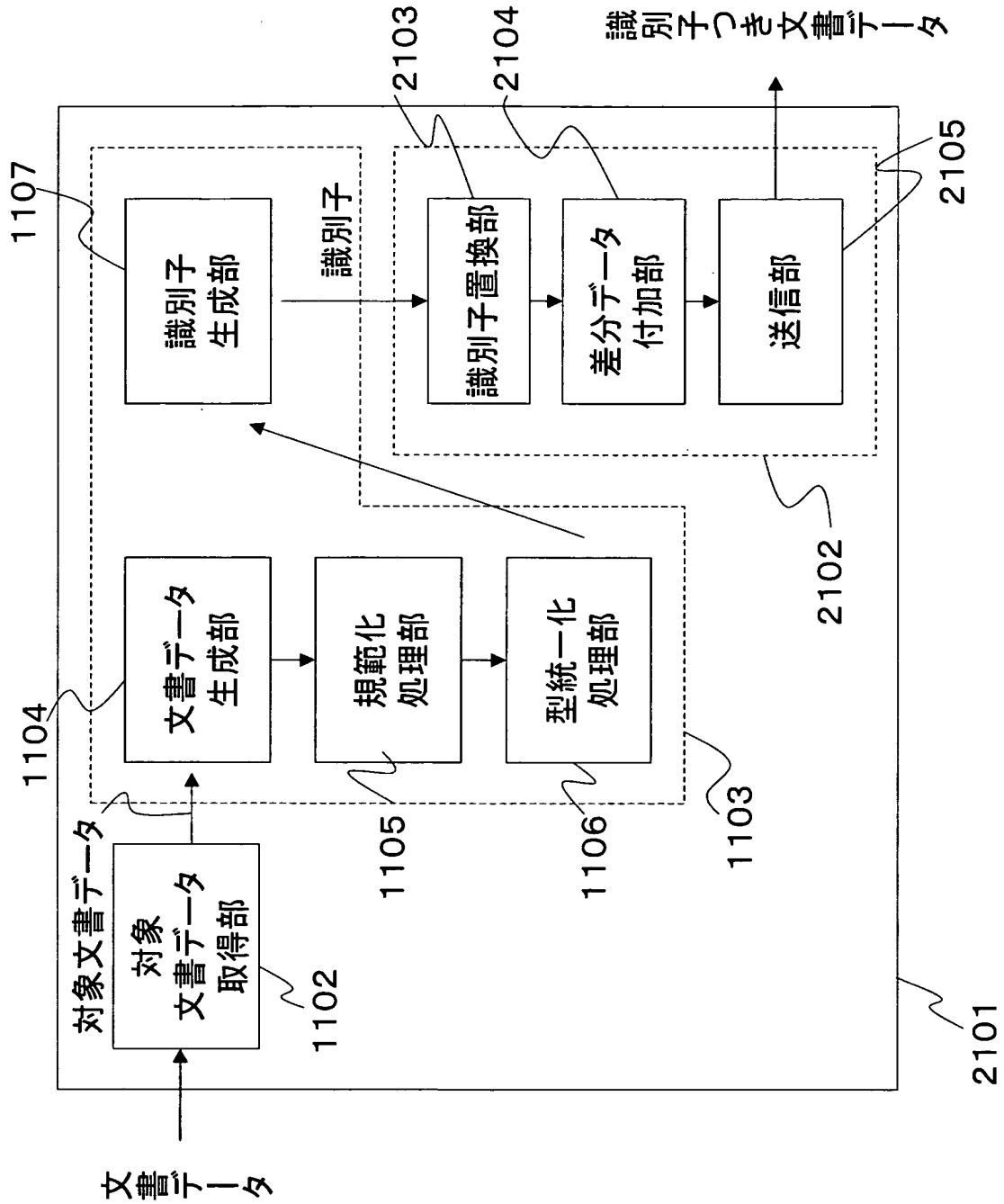
```

【図 19】

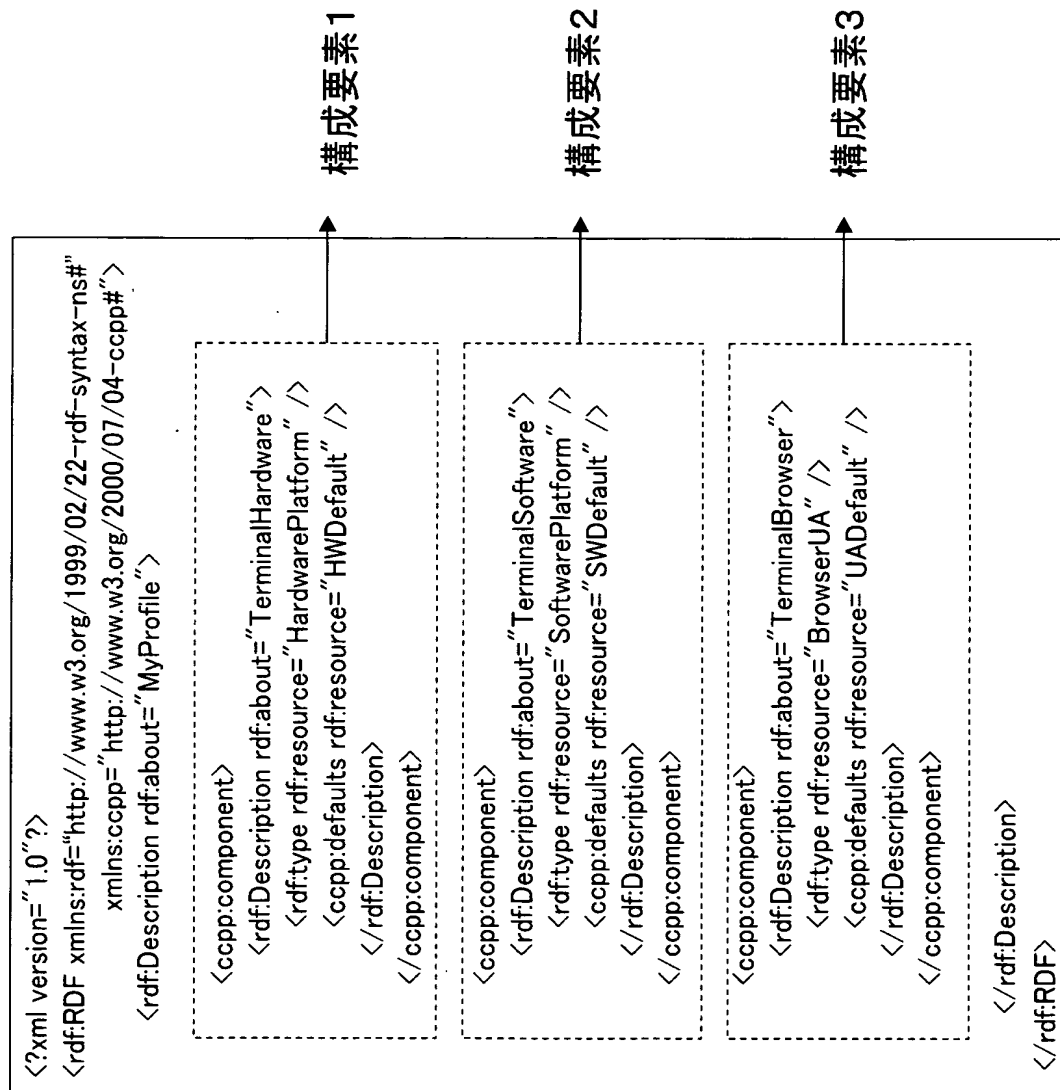
変換後文書データ

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns"
  xmlns:ccpp="http://www.w3.org/2000/07/04-ccpp"
  xmlns:prf="http://www.wapforum.org/profiles/UAPROF/ccppschem-20010330"
  xmlns:pss5="http://www.3gpp.org/profiles/PSS/ccppschem-PSS5"
  xmlns:up="http://www.nttdocomo.co.jp/profiles/UP/ccppschem-UP"
  xmlns:upf="http://www.nttdocomo.co.jp/profiles/UPF/ccppschem-UPF">
  <rdf:Description rdf:about="MyProfile">
    <ccpp:component>
      <rdf:Description ID="UserProfile">
        <rdf:type rdf:resource="UserProfilePlatform"></rdf:type>
        <up:role>vip</up:role>
        <up:age>16</up:age>
      </rdf:Description>
    </ccpp:component>
  </rdf:Description>
</rdf:RDF>
```

【図 20】



【図 21】



【図 22】

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
          xmlns:ccpp="http://www.w3.org/2000/07/04-ccpp#">
  <rdf:Description rdf:about="MyProfile">

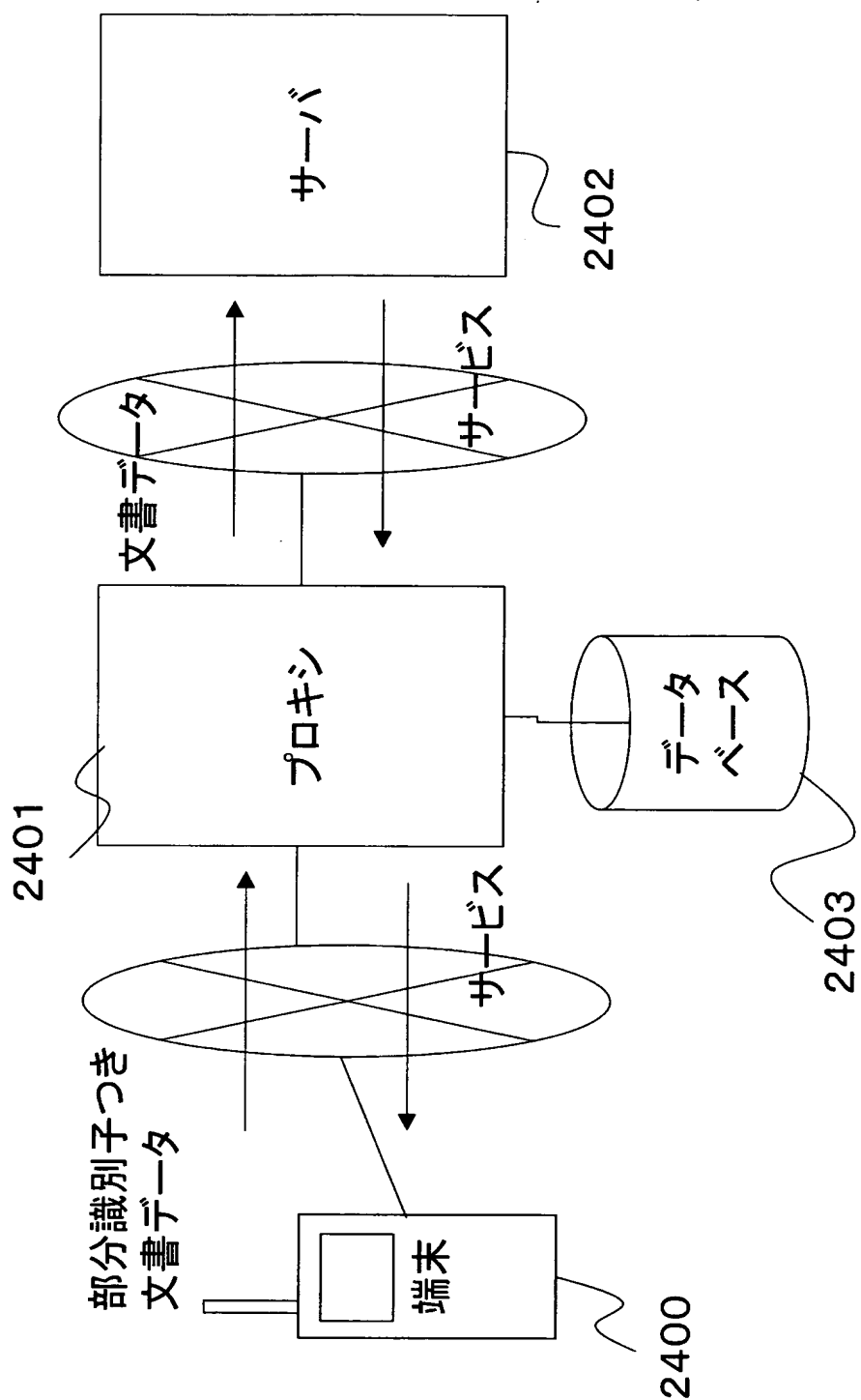
    <ref id="124515146146"/>

    <ref id="145256753415"/>

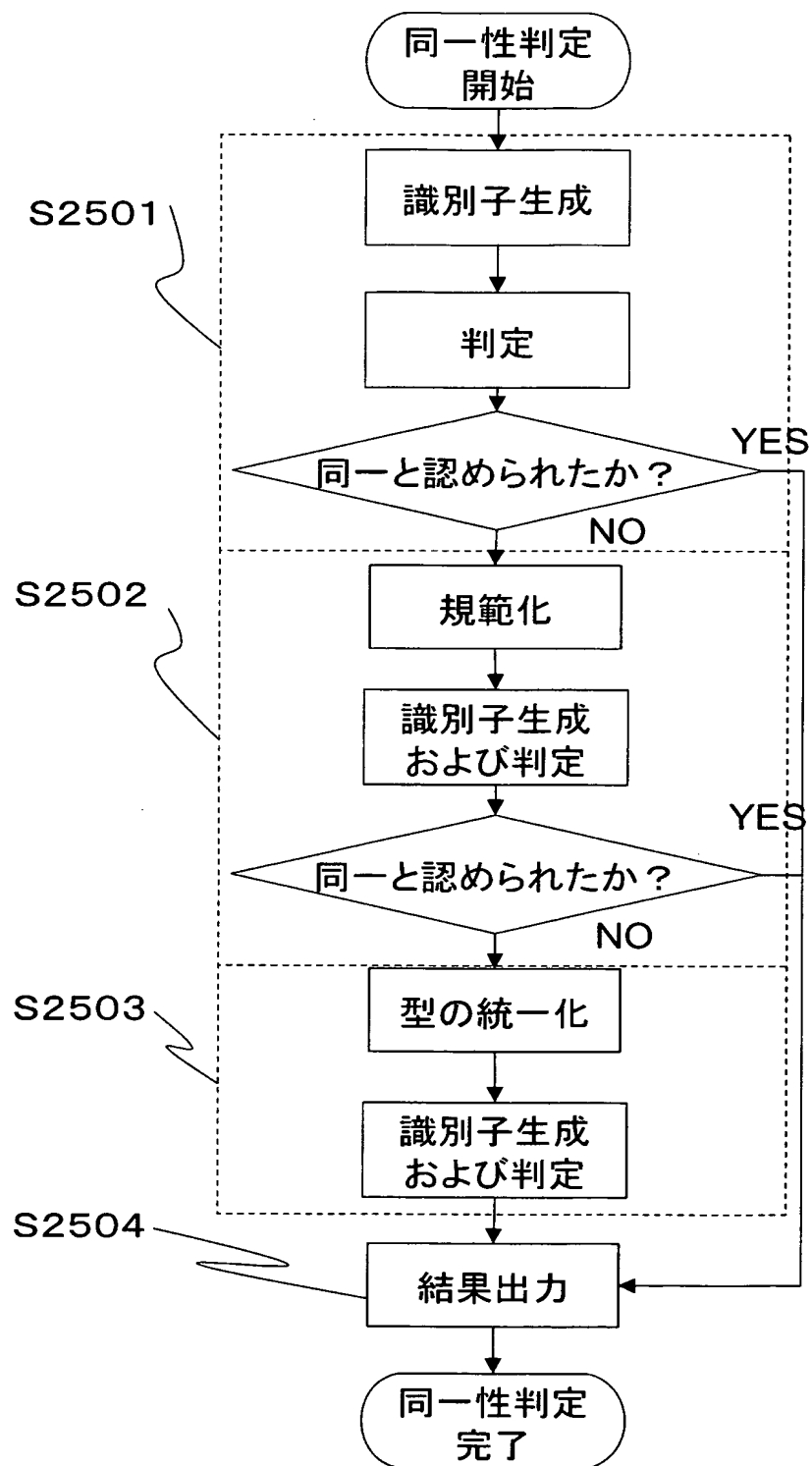
    <ccpp:component>
      <rdf:Description rdf:about="TerminalBrowser">
        <rdf:type rdf:resource="BrowserUA" />
        <ccpp:defaults rdf:resource="UADefault" />
      </rdf:Description>
    </ccpp:component>

  </rdf:Description>
</rdf:RDF>
```

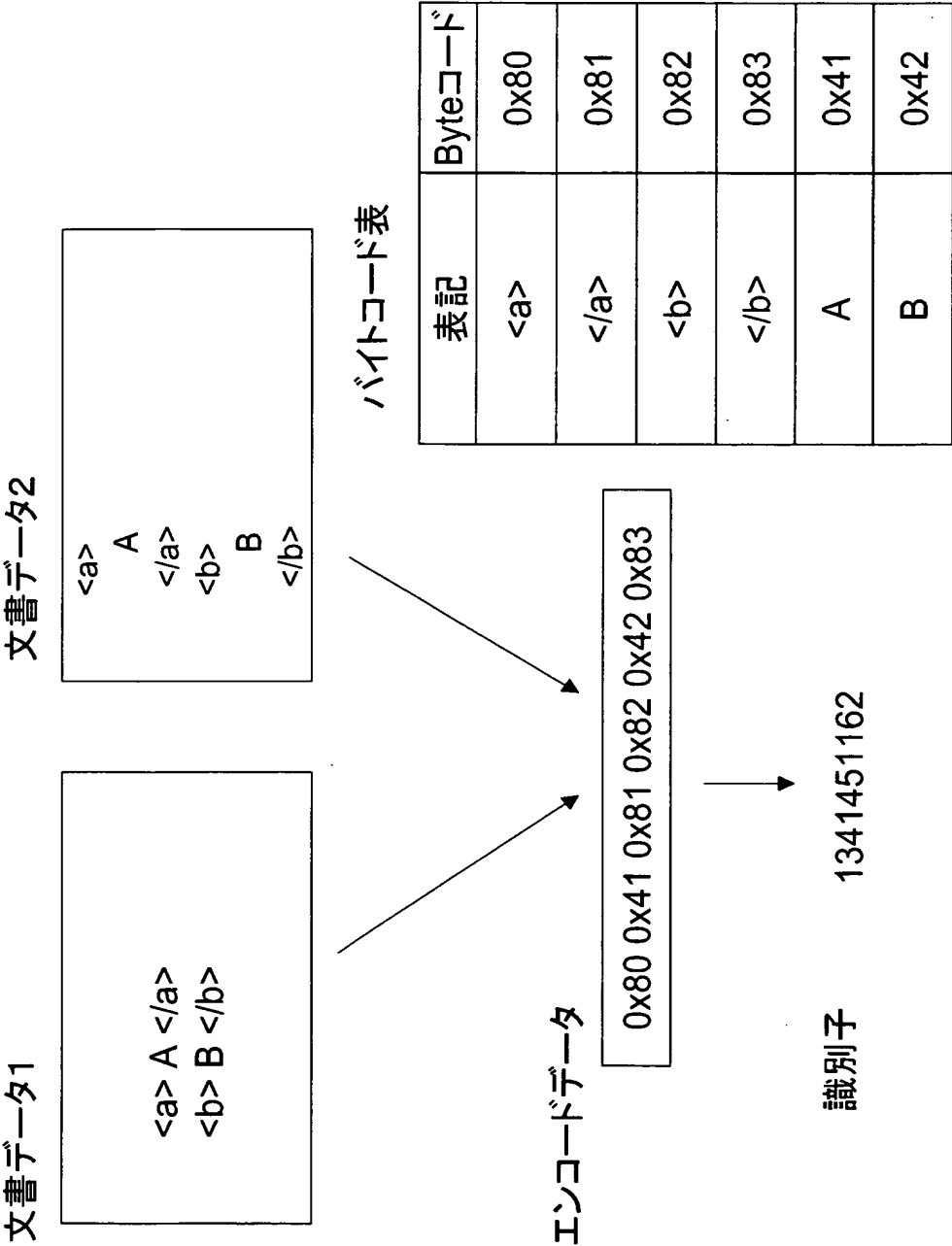
【図 23】



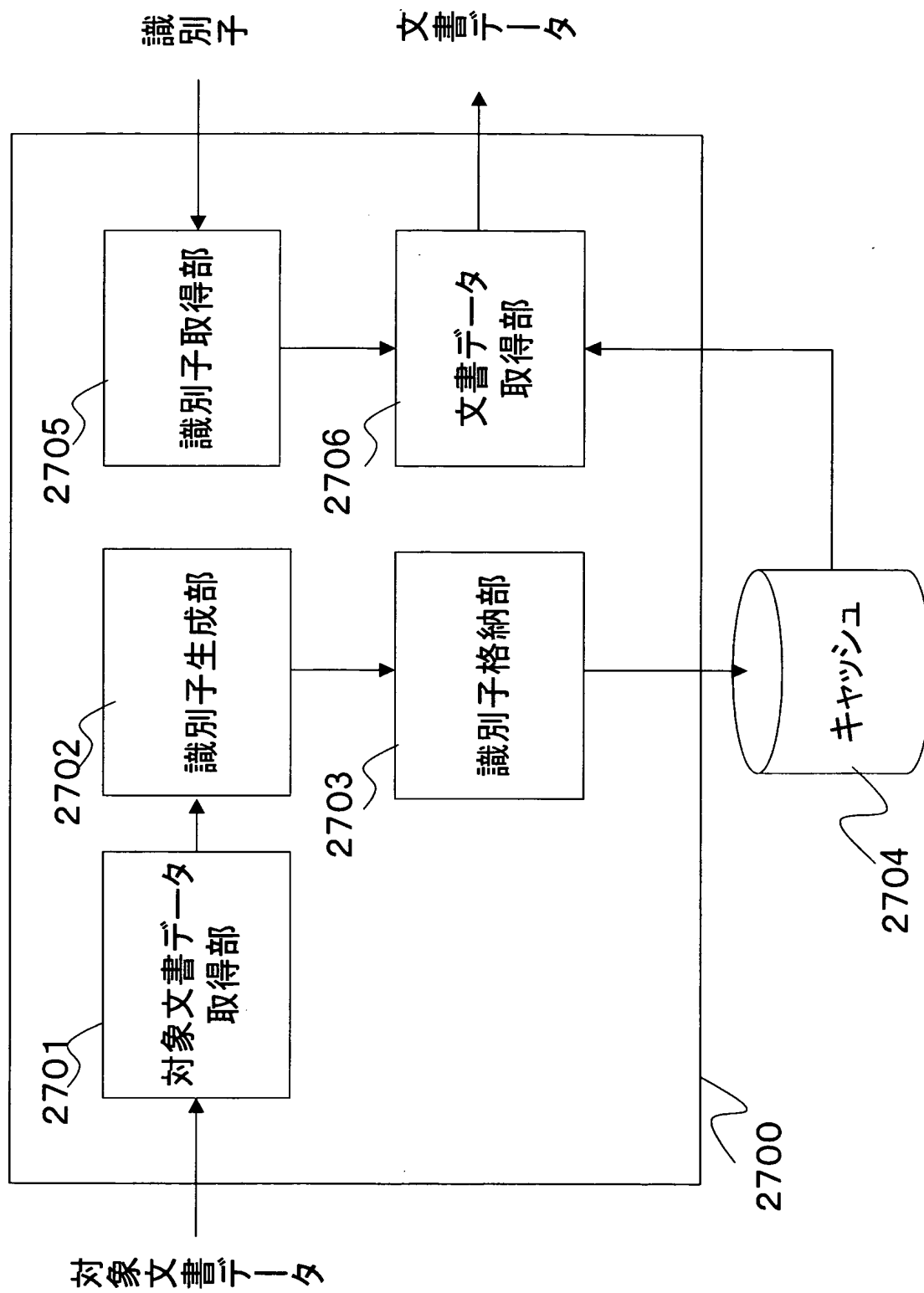
【図 24】



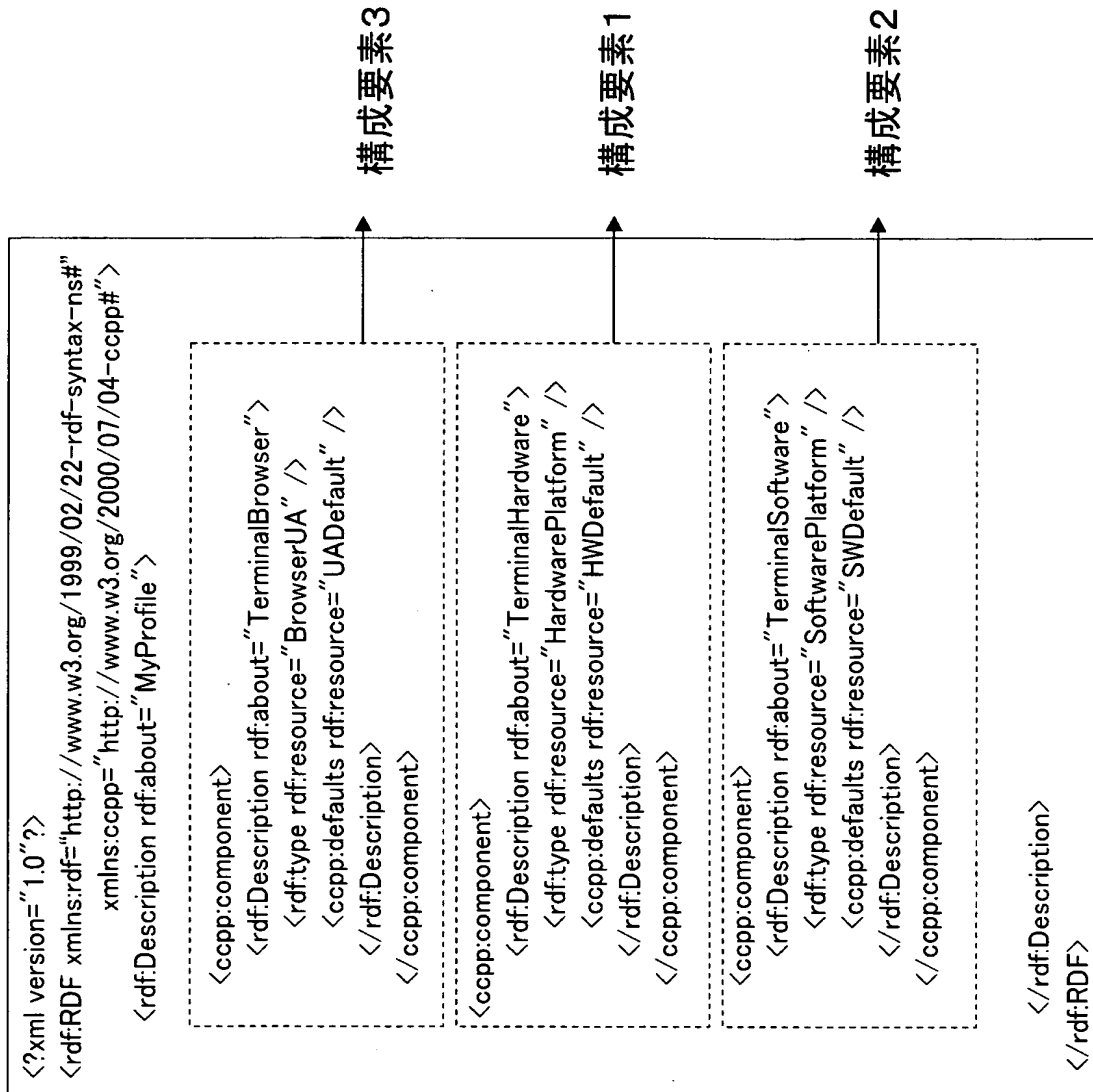
【図 2 5】



【図 26】



【図 27】



【書類名】 要約書

【要約】

【課題】 XML文書またはRDF文書が本来同一の意味を持った文書である場合、同一の識別子を生成する識別子生成方法を提供する。

【解決手段】 本発明に係る識別子生成方法は、文書データに対し、表記のゆらぎを正す規範化処理をほどこす規範化処理ステップと、規範化処理ステップにおいて規範化処理が行われた文書データ全体もしくはその一部に基づいて、当該文書データもしくはその一部を一意に特定する識別子を生成する識別子生成ステップと、を有する。このように、識別子を生成する前に規範化処理によって表記のゆらぎを回復することで、同一の意味をもつXML文書またはRDF文書は同一の表記の文書となり、一方向関数に代表される文字の羅列から識別子を生成する関数によって、同一の識別子を生成することができる。

【選択図】 図1

特願 2 0 0 2 - 2 6 7 2 7 4

出 願 人 履 歴 情 報

識別番号

[3 9 2 0 2 6 6 9 3]

1. 変更年月日

1 9 9 2 年 8 月 2 1 日

[変更理由]

新規登録

住 所

東京都港区虎ノ門二丁目 1 0 番 1 号

氏 名

エヌ・ティ・ティ 移動通信網株式会社

2. 変更年月日

2 0 0 0 年 5 月 1 9 日

[変更理由]

名称変更

住所変更

住 所

東京都千代田区永田町二丁目 1 1 番 1 号

氏 名

株式会社エヌ・ティ・ティ・ドコモ